Final Technical Report

MONOLITHIC OPTOELECTRONIC IMPLEMENTATION

OF NEURAL PLANES

3-18-92

Demetri Psaltis, Steven Lin, Annette Grot

Submitted to:

Dr. Barbara L. Yoon

Defense Advanced Research Projects Agency

3701 North Fairfax Drive

Arlington, Virginia   22203-1714

Principal Investigator

Demetri Psaltis

Department of Electrical Engineering

California Institute of Technology

Pasadena, CA   91125

Introduction

The implementation of a neural computing system, whose structure and function is motivated by natural intelligence, will provide a unique way to solve problems that are typically too difficult for the conventional electronic computers to tackle. Interest in this type of computer has emerged largely because it is hoped that by building a computer that shares some of the characteristics of the biological systems, we will be able to address problems such as image recognition which animals do exceedingly well but current computers do not. There has been considerable progress on the theoretical research in neural network to justify the optimism of future applications. This has resulted in a focused attention on the hardware realization of neural architectures. The computational power of neural computers arises from matching the computer architecture and the physical properties of the devices used in the implementation to the requirements of the problem. In other words, a neural computer is highly specialized and it is therefore very difficult to derive its full advantages on a general purpose computer. This provides a strong impetus for advancing the technologies for the physical realization of neural computers in parallel and interactively with the development of theoretical neural network models.

A neural network consists of two basic components : a large collection of neurons and a dense network of interconnections among all the neurons. Neurons are usually modeled as thresholding elements. Information is stored in the weights of the connections largely through error driven learning algorithm. If, during a learning phase, the response of the network is correct, then the connections remain unaffected. Otherwise they are modified to eventually produce a desired response. There are two contenders for the physical implementation of a neural network : elec-
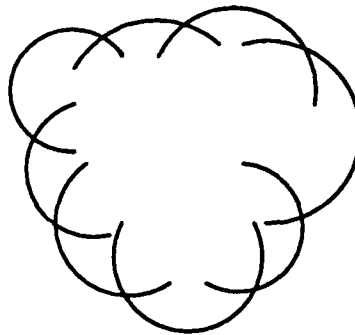
tronics and optics. While the thresholding function of a neuron is relatively easy to implement in electronics, the massively dense interconnection network among the neurons is becoming the bottleneck in the realization of an electronic neural network. Furthermore, these interconnections are not dynamically modifiable because the interconnections are defined by metal wires in the integrated circuits, which form the basic building block for the neurons. Heat dissipation and interconnection delay are also serious limiting factors as the network gets denser and larger.

Optics, on the other hand, is well suited for a system in which a network of massively interconnected elements are required. This is achieved by arranging arrays of neurons in a planar geometry and using the third dimension to globally interconnect the neural planes with light. Figure 1.1 illustrates the schematic diagram of such a system. The feature of the optical implementation that gives it an advantage when compared to the electronic counterpart is the fact that it is constructed in three dimensions. This allows the active devices at the neural planes to be populated by processing elements only, since the interconnections are external to the planes of neurons. The third dimension is used to store the information that is required to specify the connections among the neurons. It is important to keep in mind that in a densely interconnected network the weights represent a large database. This large database can be easily implemented in the form of holographic interconnections [1]. A second advantageous feature of the optical implementation is the relative ease with which learning can be accomplished by dynamic holograms recorded in photorefractive crystals [2,3]. This has allowed the holograms to be programmed in real-time and the specific interconnections to be modified as the network is in its learning phase.

The implementation of an optical network also requires physical devices that simulate the function of actual biological neurons. There are several possible candidates for the realization of optical neurons. The first is an optically addressed
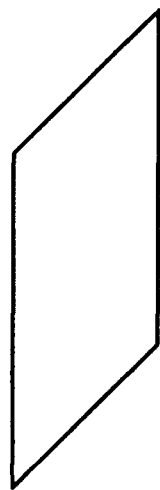
Plane of

Input Neurons

Optically

Implemented

Interconnections

Plane of

Output Neurons

**Fig. 1.1** Architecture of an optically implemented neural network.

spatial light modulator such as the liquid crystal light valve [4]. Though large in density, liquid crystal light valves are not flexible in their use. The lack of variable threshold control and resolution also contribute to their functional inadequacy. In addition, the temporal response is in the milli-second range, which is slow for some applications. Other devices like ferroelectric liquid crystal spatial light modulators on silicon [5], electrooptic ceramics, such as PLZT, on silicon [6], heteroepitaxy of III-V material on silicon [7], and epitaxial lift-off hybridization of fabricated III-V devices on silicon [8] have also been tried. These devices are taking the advantage of the relatively mature silicon VLSI integrated circuits technology to provide the necessary functionality. However, to have an optical output, either light emitters, such as GaAs lasers or LED's, or spatial light modulators are either built somewhere else on different material and then transported to silicon circuits or are grown on the silicon substrate directly. The former method, which involves hybridization of two incompatible devices, requires complicated processes and procedures in properly connecting the optical devices to the silicon circuits. The latter method involves growing typically GaAs on silicon substrate. Because of the build-in 6% lattice mismatch in the lattice spacing between GaAs and Si, monolithically integrated devices based on GaAs on Si are subject to strain, which, if improperly controlled, will result in defects in the material and cause the degradation in the device performance.

The third candidate for the realization of the optical neurons is monolithically integrated optoelectronic circuits [9]. It can provide a better solution much faster and much easier. The optoelectronic approach is to construct a two dimensional array of elements with each element in the array comprised of monolithically integrated detectors, electronic amplifiers and light sources. Each element simulates an individual neuron. The entire device can be built using well established fabrication techniques in GaAs and large two dimensional arrays can be constructed. The light sensitivity is excellent when compared to either liquid crystal spatial light modula-

tors or hybridized devices. Potentially very high optical gains can also be achieved in optoelectronics to allow for the large fanout, which is required for a massively interconnected network. Moreover, there is flexibility in slowing the device down to any desirable speed in order to accommodate the large monolithic arrays to operate with reasonable electrical power dissipation. There is also flexibility in setting the function of each neuron, for example, threshold level and sharpness, electronically by designing the device and circuit appropriately. As a result, monolithically integrated optoelectronic circuit offers overall superior performance and greater flexibility when compared to the other candidates for optical neurons. Therefore, a neural network system, in which a planar array of GaAs optoelectronic integrated circuits with holographic optical elements located on top of the array to provide the interconnections among the neurons, can be envisioned. This is conceptually illustrated in Fig. 1.2.

This report discusses the results from an investigation of various optical and electronic devices which, when monolithically integrated, give the best performance for the optoelectronic neurons. This involves detailed study on various discrete devices as well as integrated devices. Because of the nature of monolithic integration, the best discrete devices may not necessarily yield the best optoelectronic neuron when they are monolithically integrated. Consequently, trade-off analysis is required to determine the relative importance of each device in the integration. Nevertheless, high-performance discrete devices are generally desirable, regardless of the level of integration involved. Some examples of such performance are high-efficiency LED's, laser diodes and photodiodes, high-current-gain bipolar transistors, and high-transconductance field-effect transistors. Some background information on bipolar transistors and field-effect transistors is included in order to fully explore the trade-off among various schemes of integration. By properly designing the circuit, low-power and high-gain optoelectronic neurons are demonstrated.
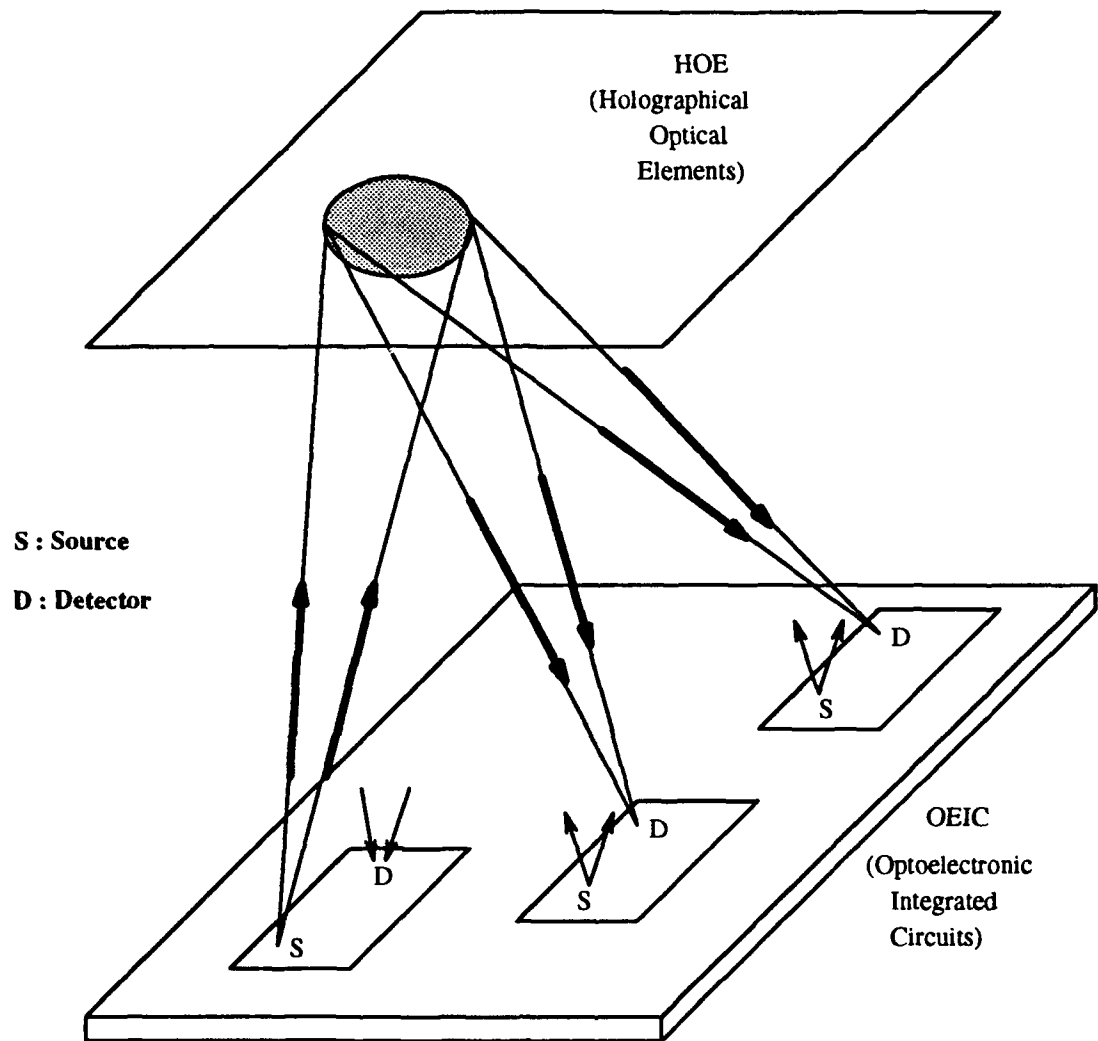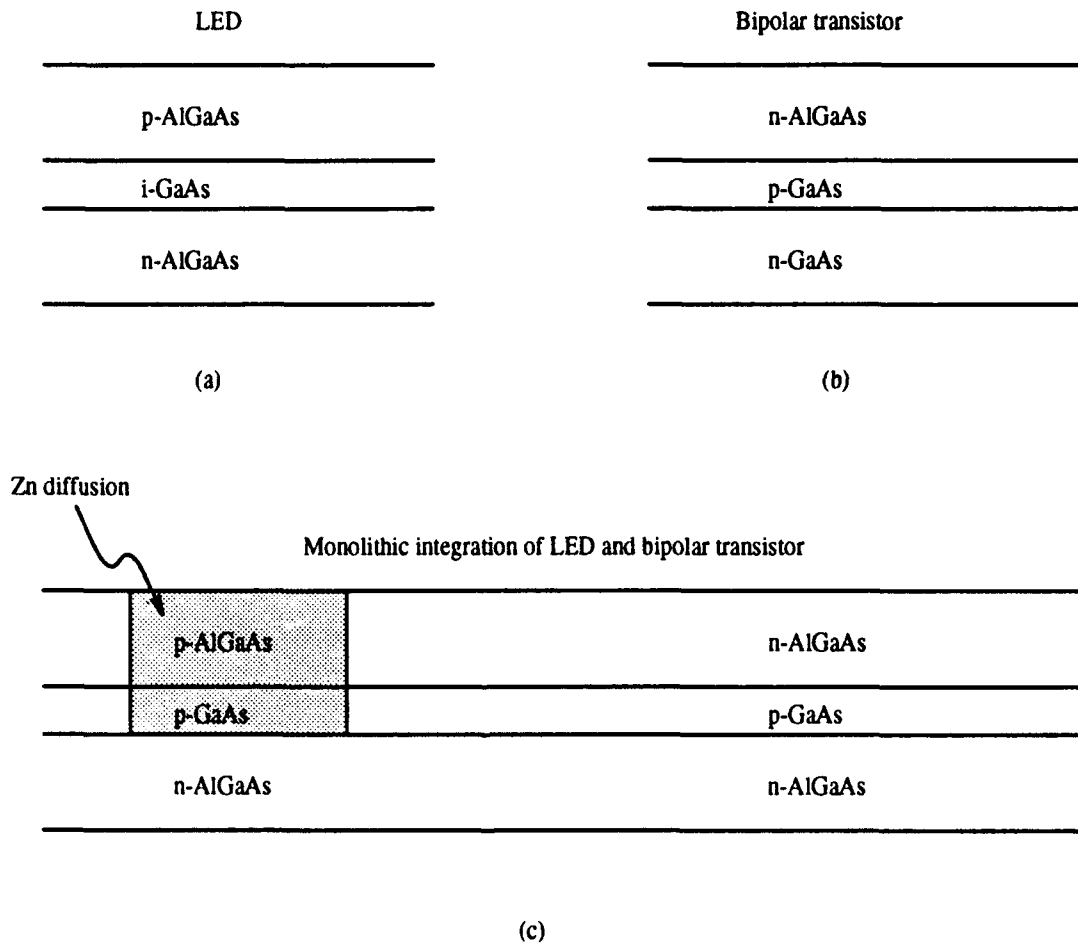
Fig. 1.2 Implementation of an optical neural network utilizing the well established GaAs optoelectronic integrated circuits as optical neurons and holographic optical elements as interconnection media.

Double-Heterojunction Bipolar Transistor-Based Neurons

2.1 Introduction

The input-output characteristics of an optoelectronic neuron is approximated by a thresholding function, whose output remains zero until the input exceeds a predetermined threshold. Beyond this threshold, the output level saturates. This input-output functionality can be easily implemented in integrated circuits fabricated by standard GaAs processing technology. Since the neuron has to have an optical input and an optical output, additional elements, such as the light detector and the light emitter, have to be added to the integrated circuits in order to obtain a complete optical neuron. Bipolar transistors are suitable for this type of integration because, in addition to the inherent gain provided by the transistors, they can function as phototransistors, which detect light with high efficiency. Thus, the need for transistor amplifiers and detectors can be simultaneously satisfied by bipolar transistors. The remaining issue is the integration of these transistors with a light emitter. For reasons of reducing overall electrical power dissipation in an array of optoelectronic neurons, light-emitting diodes (LED's) are chosen as the light emitter. Monolithic integration of bipolar transistors and LED's present a problem in the material compatibility. The structure of a LED usually consists of an active material, such as GaAs, sandwiched between two higher bandgap cladding materials, such as AlGaAs, in a double-heterostructure fashion. The doping requirement for the LED is heavy p^+ doping in the top AlGaAs cladding layer, intrinsic doping in the GaAs active layer and heavy n^+ doping in the bottom AlGaAs cladding layer, forming a P-i-N diode. This is illustrated in Fig. 2.1(a). For the bipolar transistor, however, the structure is somewhat different. The emitter is usually a high bandgap material, such as AlGaAs, and the base and the collector are low bandgap material,
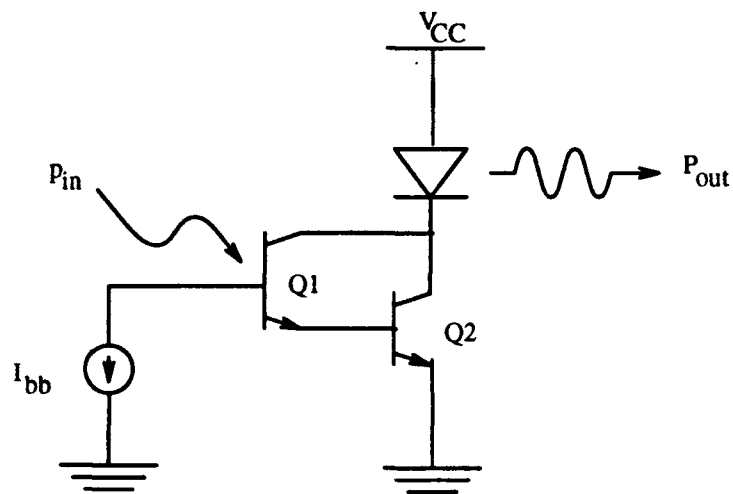
LED	Bipolar transistor
p-AlGaAs	n-AlGaAs
i-GaAs	p-GaAs
n-AlGaAs	n-GaAs
(a)	(b)

Zn diffusion

Monolithic integration of LED and bipolar transistor

p-AlGaAs	n-AlGaAs
p-GaAs	p-GaAs
n-AlGaAs	n-AlGaAs

(c)

Fig. 2.1 Structure of typical epitaxial layers for **(a)** LED and **(b)** heterojunction bipolar transistor. By converting the collector to a higher bandgap material in the transistor and n-type upper cladding layer to p-type in the LED, both the LED and the bipolar transistor can be fabricated in the same epitaxial layer as shown in **(c)**.

such as GaAs. The doping composition in a bipolar transistor is n-type, p-type and n-type in emitter, base, and collector respectively in order to utilize the high electron mobility in the base. This is illustrated in Fig. 2.1(b). As seen in the figure, the material and doping requirements for the LED and the bipolar transistor are somewhat different from each other. For the top layer, which is AlGaAs for both device, LED is doped p-type and the transistor is doped n-typed. For the second layer, which is GaAs in both devices, the LED is intrinsic and the transistor is p-type. For the last layer, both devices are doped n-type. However, the LED requires a high bandgap material whereas the transistor usually has a small bandgap material. In order to successfully integrate these two devices in a planar fashion, some compromise from each device has to be made. For example, the collector of the bipolar transistor does not have to be GaAs. It can be AlGaAs, which will make the transistor a double-heterojunction bipolar transistor (DHBT) and matches the lower cladding AlGaAs in the LED. In addition, the active layer of the LED can be doped p-type as long as it is sandwiched between two large bandgap materials. For the top layer, a compromise can not be made easily because each device requires a totally opposite doping composition from each other. Thus, a conversion from n- to p-type or vice versa has to be performed in order to obtain both devices simultaneously. While n-type diffusion is harder to perform and is not characterized as well, p-type diffusion can be performed in a very controlled manner on the LED to convert the originally n-type AlGaAs to the p-AlGaAs, which is required for the LED. This is graphically depicted in Fig. 2.1(c). In fact, Katz. et al. has experimentally demonstrated the feasibility of fabricating both the bipolar transistor and the LED from the same epitaxial material by performing Be implantation to convert the n-type AlGaAs to p-AlGaAs [10]. This structure will form the basic material structure to build the DHBT-based optoelectronic neurons.
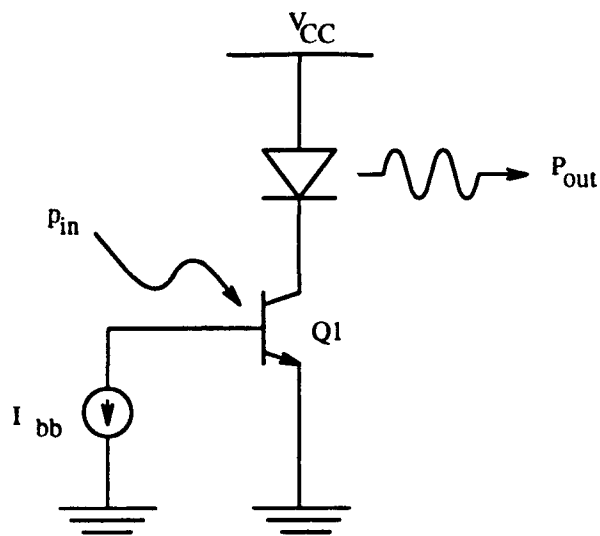
(a)

(b)

Fig. 2.2 (a) Schematic circuit diagram of an optoelectronic neuron incorporating only one bipolar transistor. (b) Schematic circuit diagram of an optoelectronic neuron incorporating two bipolar transistors to provide the gain needed to satisfy the loop gain requirement.

2.2 Design Considerations

The integration of LED's with heterojunction bipolar transistors presents a unique approach to realizing optoelectronic neurons needed for the neural network implementation. Shown in Fig. 2.2(a) is the schematic circuit diagram of the optoelectronic neuron that incorporates both devices. The bipolar transistor not only functions as the amplifier, but also as the photodetector. The thresholding is provided by applying a reverse biased current, I_{bb}, on the base of the transistor such that the transistor will not be turned on until the photogenerated current has exceeded the reverse biased current. After which, the transistor amplifies the signal received to produce an output current that drives the LED. This process continues until the transistor saturates, which causes the neuron to saturate as well. In order for this circuit to work in a neural network properly, several issues have to be addressed. Firstly, this optoelectronic neuron has to be able to provide sufficient optical gain in order for the signal to propagate to the next neuron without dying down. This implies that high current gain from the bipolar transistor is a requirement. If we assume η_H, η_D, η_L, and β are the efficiencies of the hologram that specifies the interconnections, the LED, the detector, and the current gain of the bipolar transistor respectively, then it is necessary to mandate the following relationship in order to close the loop without any attenuation :

$$\eta_H \cdot \eta_D \cdot \eta_L \cdot \beta \geq 1. \tag{2.1}$$

For $\eta_H = 0.1$, $\eta_D = 0.3 A/W$, and $\eta_L = 0.01 W/A$, β has to be at least 3333. Though a current gain of greater than 5000 has been reported in GaAs heterojunction bipolar transistors [11,12], it may be difficult to fabricate transistors that satisfy this gain

reliably and consistently. Thus, two heterojunction bipolar transistors connected in a Darlington pair configuration has been proposed to meet the current gain requirement for the neuron and at the same time provide a reliable and practical way of fabricating the transistors. This is shown in Fig. 2.2(b). By using a Darlington pair, a combined current gain of 3333 can be obtained more easily as the product of the current gains from each transistor, $\beta_1 \cdot \beta_2$, need only be greater than 3333.

The second issue of concern is the ability of these bipolar transistors to provide gains at low driving power. It is well-known that the current gain of the transistor is dependent upon the collector current. In fact, the higher the current gain required, the higher the collector current needed, which, in turn, increases the power dissipation of the transistors. Approximately, the relationship can be expressed as

$$\beta \sim I_c^{1-\frac{1}{n}}, \tag{2.2}$$

where n is the ideality factor for the base-emitter junction, which ranges from 1 for the ideal junction to 2 for the non-ideal junction. The case of n = 2 corresponds to the situation in which the base current is dominated by recombination taking place through deep level traps in the space charge region. If the base-emitter junction is ideal, it can be seen from Eq. (2.2) that β is independent of the collector current. However, if the junction is not ideal, β's dependence on I_c can be as dramatic as square root. Thus, to achieve the high gain needed by the neuron, it is not surprising if the level of collector current needed is higher. To circumvent this problem, we need to decrease the ideality factor to as close to one as possible. In other words, the current component that contributes to the ideality factor of two should be minimized. This can be achieved by designing the base-emitter junction such that the depletion region is narrow enough to disallow recombination within this region. Therefore, high β's can be attained at low collector currents.

The third issue is on the performance compromise of the LED and the transistor as a result of sharing the same epitaxial layers required for monolithic integration. This compromise arises from the fact that the p-GaAs is shared by all three devices. For the photodetector, this p-GaAs is the absorption layer, which needs to be thick to allow for the complete absorption of incoming photons. However, for the transistor, this layer is the base, which should be as thin as possible to maximize the current gain. Similarly, for the LED, it can not be too thick or too thin due to self-reabsorption and interfacial recombination. Thus, the thickness of the p-GaAs layer needs to be carefully chosen so that the overall performance of the neuron, not each individual device, is maximized.

To quantify the parameters of the transistors before monolithically integrating them with the LED on the same substrate, individual transistors were first characterized to ensure the β's measured was sufficient for the Darlington transistor pair to provide the current gain needed for the neurons.

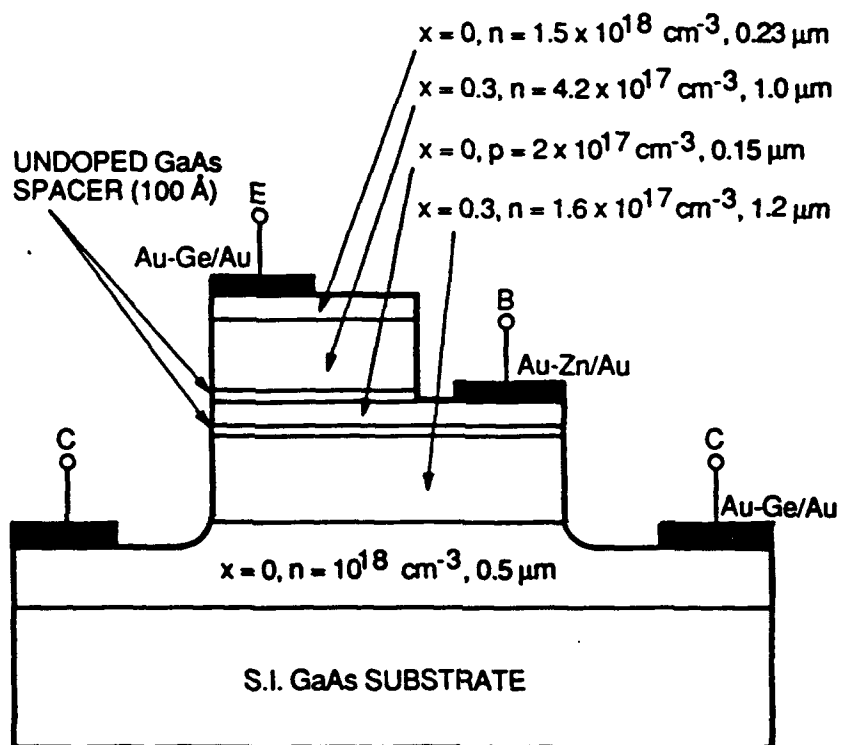## 2.3 Discrete Double-Heterojunction Bipolar Transistors

GaAlAs/GaAs/GaAlAs double heterojunction bipolar transistors (DHBT's) are very attractive for high-gain applications and optoelectronic monolithic integration because of their structural compatibility with laser diodes [10,13] and LED's. Very high current gain ($\beta \sim 10^4$) has been demonstrated in single heterojunction bipolar transistors (SHBT's) grown by liquid phase epitaxy [12,14]. However, most of the SHBT's grown by molecular beam epitaxy (MBE) [15] and MOCVD [16] show much lower gains. The current gain is even lower in DHBT's [11,17]. In spite of some encouraging results [18,19] and recent progress in the crystal growth by MBE and MOCVD, the reproducibility of high-gain heterojunction bipolar transistors, especially DHBT's, is not sufficiently good mainly because the heavy base doping

concentration incorporated in these transistors has resulted in an out-diffusion of these base dopants during crystal growth or subsequent high-temperature processing. By introducing spacer layers at the emitter-base junction and the collector-base junction and reducing the doping in the base layer, the diffusion of the base dopants can be minimized.

The double heterojunction structure was epitaxially grown on (100) Cr-doped semi-insulating GaAs substrates ($\rho \geq 5 \times 10^7$ ohm·cm) by a metalorganic chemical vapor deposition system (SPIRE-450) using a vertical barrel reactor. The GaAs and AlGaAs layers were grown by trimethyl gallium (TMG), trimethyl aluminum (TMA), and 10% AsH_3 in 90% H_2. Zinc and silicon were used for p- and n-type dopants, respectively. The substrate temperature during the growth was about 730° C. DHBT's structure consists of : 0.5 μm of Si-doped (10^{18} cm^{-3}) n-GaAs subcollector/buffer, 1.2 μm of Si-doped (1.6×10^{17} cm^{-3}) n-$Al_{0.3}Ga_{0.7}As$ collector, 100$\mathring{A}$ of undoped GaAs spacer layer, 0.15 μm of Zn-doped (2×10^{17} cm^{-3}) p-GaAs base, 100 $\mathring{A}$ of GaAs undoped spacer layer, 1.1 μm of Si-doped (4.2×10^{17} cm^{-3}) n-$Al_{0.3}Ga_{0.7}As$ emitter, and 0.23 μm of Si-doped (1.5×10^{18} cm^{-3}) n-GaAs cap layer.

The structure of the DHBT is schematically shown in Fig. 2.3. The devices with an emitter area of 2.4×10^{-4} cm^2 were fabricated by standard lift-off and wet chemical etching processes. The base was properly exposed by first etching the GaAs cap layer in H_3PO_4 + H_2O_2 + CH_3COOH, followed by etching the $Al_{0.3}Ga_{0.7}As$ emitter in NH_4OH + H_2O_2 + H_2O. H_3PO_4 + H_2O_2 + CH_3COOH was used to etch epilayers down to the subcollector layer for collector contacts. AuGe/Au and AuZn/Au were evaporated for emitter/collector and base contacts, respectively and alloyed separately.

Typical common-emitter current characteristics are shown in Fig. 2.4 at several different current levels. Current gains of 40, 100, 300, and 500, which was the highest

Fig. 2.3 The schematic cross sectional view of the $Al_xGa_{1-x}As/GaAs/Al_xGa_{1-x}As$ double heterojunction bipolar transistor.

current gain reported for MOCVD-grown DHBT's without base or junction grading then, were obtained at collector currents of 0.2, 10, 70, and 120 mA, respectively as shown in Fig. 2.4(a), (b), (c), and (d). The collector current density at which the current gain of 500 was obtained was 500 A/cm^2 based on the base-emitter junction area of 2.4x10^{-4} cm^2. Figure 2.4(d) also shows the inverted-mode DHBT characteristics with a current gain of 10 at an emitter current of 2 mA. The forward I-V characteristics showed an offset voltage of 0.3 V. However, no offset voltage was observed in the inverted mode. Reverse breakdown voltages of 8 and 10 volts were observed for the emitter-base and the collector-base junctions, respectively. The collector-emitter breakdown voltage (BV_{CEO}) in the common-emitter configuration was 5 V.

Figure 2.5 shows the logarithmic plot of the measured common-emitter current gain as a function of the collector current. The ideality factor evaluated from the relation of $\beta \sim I_c^{1-\frac{1}{n}}$ was approximately 1.2. This value indicates that the recombination current in the emitter-base junction depletion region is not negligible at small collector currents. It should be noted however that the current gain increases continuously with increasing collector current. This means that no serious base push-out or emitter crowding effect exists. In this device operating region, a maximum current gain of 500 was obtained. We observed a higher current gain of 750 at a higher collector current. However, DHBT's operated at this current level were not thermally stable. These high current gains are evidence of effective blockage of Zn out-diffusion as a result of the reduced base doping concentration and the insertion of the undoped GaAs spacer layers. However, these inserted undoped spacer layers may be responsible for not having a smaller base-emitter junction ideality factor of 1.4 because of the recombination current taking place within these depleted regions.

By connecting two of these transistors in a Darlington fashion, a combined current gain of 4000 has been measured. This is shown in Fig. 2.6. Because of the
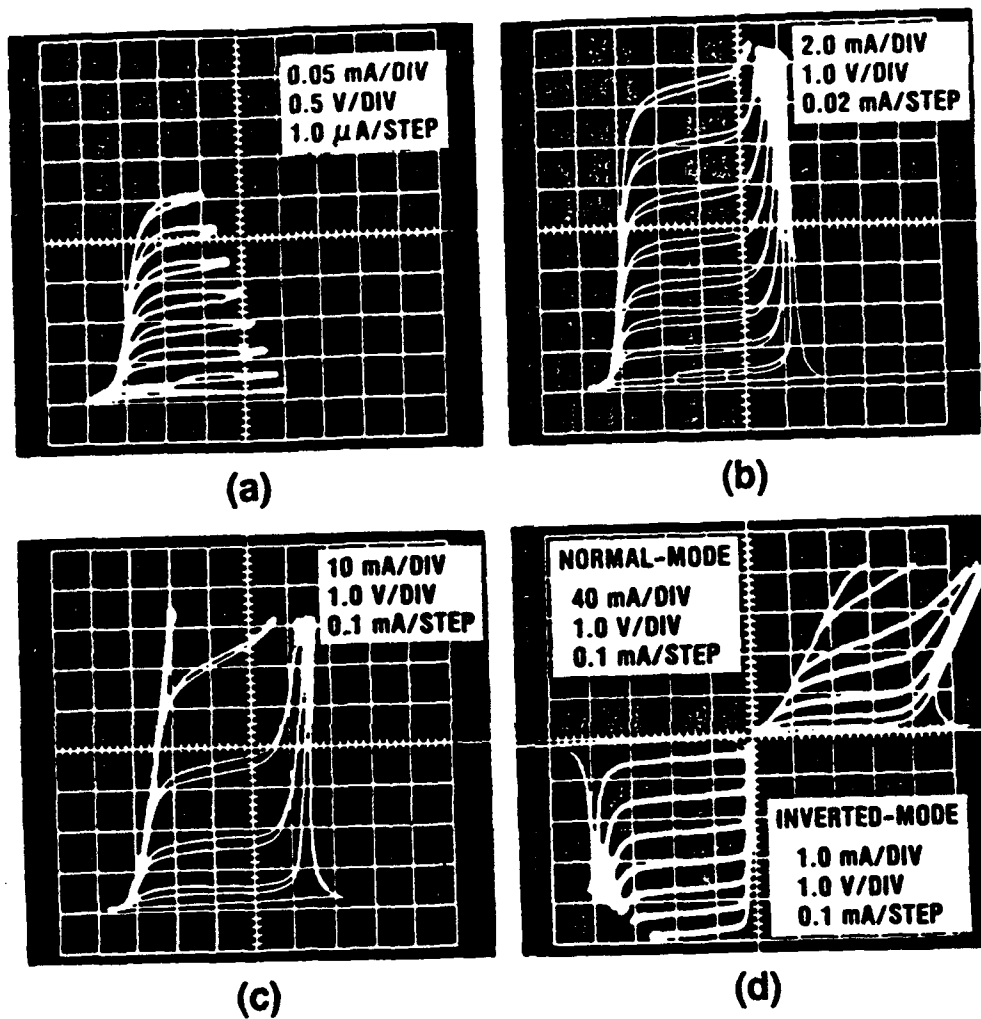
Fig. 2.4 Typical common-emitter I-V characteristics of the DHBT at: **(a)** low current levels; **(b)** normal current levels; **(c)** high current levels; **(d)** normal and inverted modes.
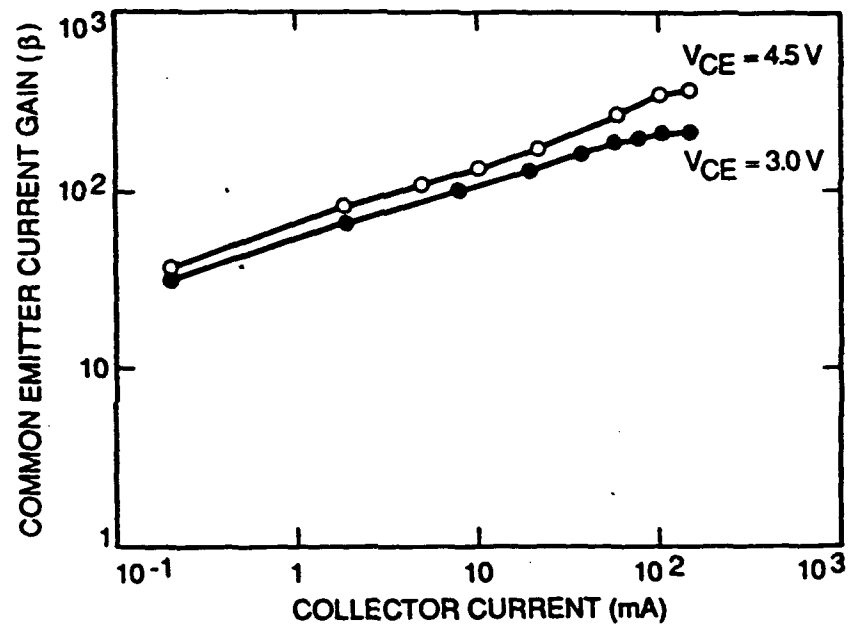
Fig. 2.5 The common-emitter current gain as a function of the collector current at $V_{CE} = 3$ and 4.5 V.

electrical contact problem, a large offset voltage in V_{CE} was observed. Nevertheless, it showed the feasibility of achieving the current gain required for the integrated optoelectronic neuron by using a Darlington transistor pair.

2.4 Monolithically Integrated Optoelectronic Neurons

With the current gain of 500 demonstrated in the discrete double-heterojunction bipolar transistors, monolithically integrated optoelectronic neurons consisting of two double-heterojunction bipolar transistors with a LED on a common GaAs substrate as shown in the circuit in Fig. 2.2(b) are next fabricated. Since the design and material parameters for the discrete transistors are approximately the same as those of the integrated Darlington transistor pair, the integrated Darlington transistor pair is expected to exhibit the same high current gain as that observed in discrete transistors. However, a minor difference exists between the discrete and the integrated transistor. In the discrete transistor, base contacts are defined by etching down to expose the base, followed by evaporation of proper metals. For the integrated transistor, Zn-diffusion is used to facilitate the making of the contact to the base of the transistor. This avoids the need for etching down to expose the base, which is a very sensitive and delicate process. This is because the base layer is so thin that it is very easy to over-etch it. While Zn-diffusion is performed to make contact to the base of the transistor, it also serves as a necessary step in converting the n-AlGaAs upper cladding layer to p-AlGaAs, thus forming a P-i-N diode for the LED. In fact, the reason why Zn-diffusion is chosen to make contact to the base of the transistor is because the n-AlGaAs upper cladding layer needs to be Zn-diffused in converting to p-type AlGaAs anyway. Thus, while this Zn-diffusion is necessary for formation of the LED, it also facilitates making the contact to the base of the transistor so that one extra step in the processing of this
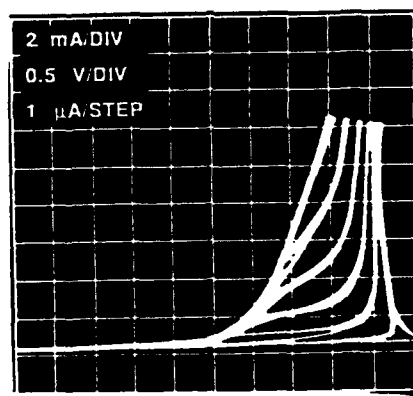
Fig. 2.6 The common-emitter I-V characteristics of two discrete DHBTs' connected a Darlington pair fashion. A combined current gain of 4000 has been measured. An offset voltage of 2.5 V in V_{CE} was observed because of the electrical contact problem of the probes.
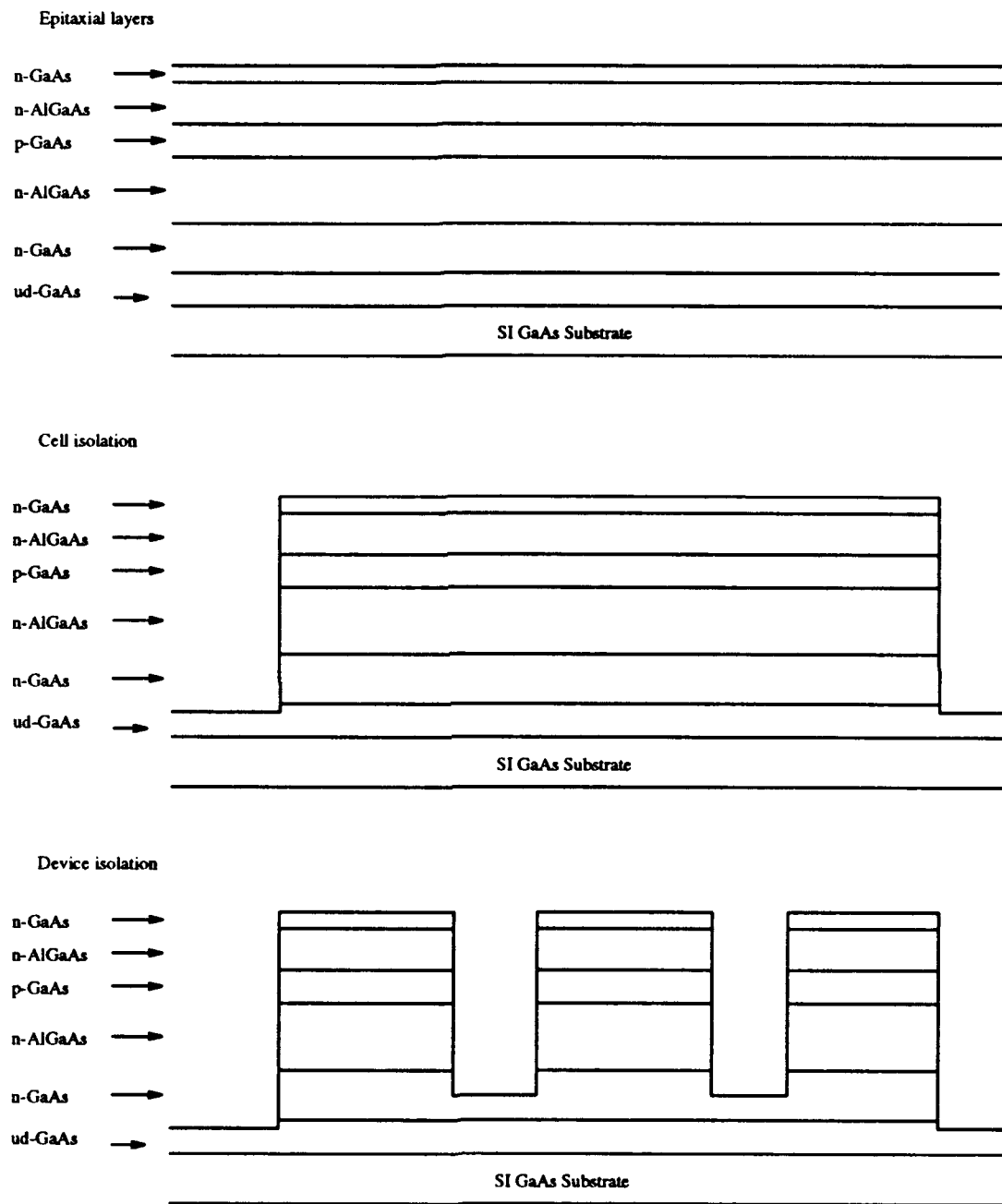
integrated optoelectronic neuron can be eliminated.

Figure 2.7 shows the cross sectional view of the optoelectronic neuron. The structure of the epitaxial layers is the same as that for the discrete transistors and is described earlier in Sec. 2.3. Figure 2.8(a) through 2.8(c) illustrate the step-by-step process in fabricating the optoelectronic neurons. Following the standard post-growth wafer cleaning procedure, each neuron was first photolithographically defined by etching the epilayers into the semi-insulating substrate with a nonselective etchant, $H_3PO_4 + H_2O_2 + CH_3COOH$. Each individual device in a neuron was subsequently defined by etching the epilayers into the AlGaAs collector layer using the same etchant so that the DHPT, DHBT and LED were mutually connected by the n-type GaAs sub-collector layer only. Zn-diffusion at 650°C for 45 minutes was then performed in a sealed ampoule using $ZnAs_2$ as the source in order to provide external base contacts for the Darlington transistor pair as well as to convert the n-type cap and emitter layers to p-type for the LED. The mask used for diffusion was Si_3N_4 grown at 680°C by a thermal chemical vapor deposition system. The diffusion regions were defined by etching the Si_3N_4 mask in a CF_4 plasma. Following Zn-diffusion, the photosensitive area of the DHPT was opened by removing Si_3N_4. Cr/Au was evaporated and lifted off for both the p-type ohmic contacts and the interconnection lines. AuGe/Au was evaporated and lifted off for the n-type ohmic contacts and alloyed subsequently at 380° for 1 minute. Each array has dimensions of 5×5 mm^2 and each neuron has dimensions of 250×250 μm^2. The light-emitting area of the LED is 8×8 μm^2 and the light-detecting area of the DHPT is 50×130 μm^2. The emitter areas for the DHBT and the DHPT are 3.5×10^{-5} cm^2 and 1.6×10^{-4} cm^2 respectively.

A 10×10 array of the optoelectronic neurons was also fabricated by using the same process procedure. It has dimensions of 5mm $\times$ 5mm with 40 bond pads surrounding the array. These 100 neurons were grouped in a certain fashion so that
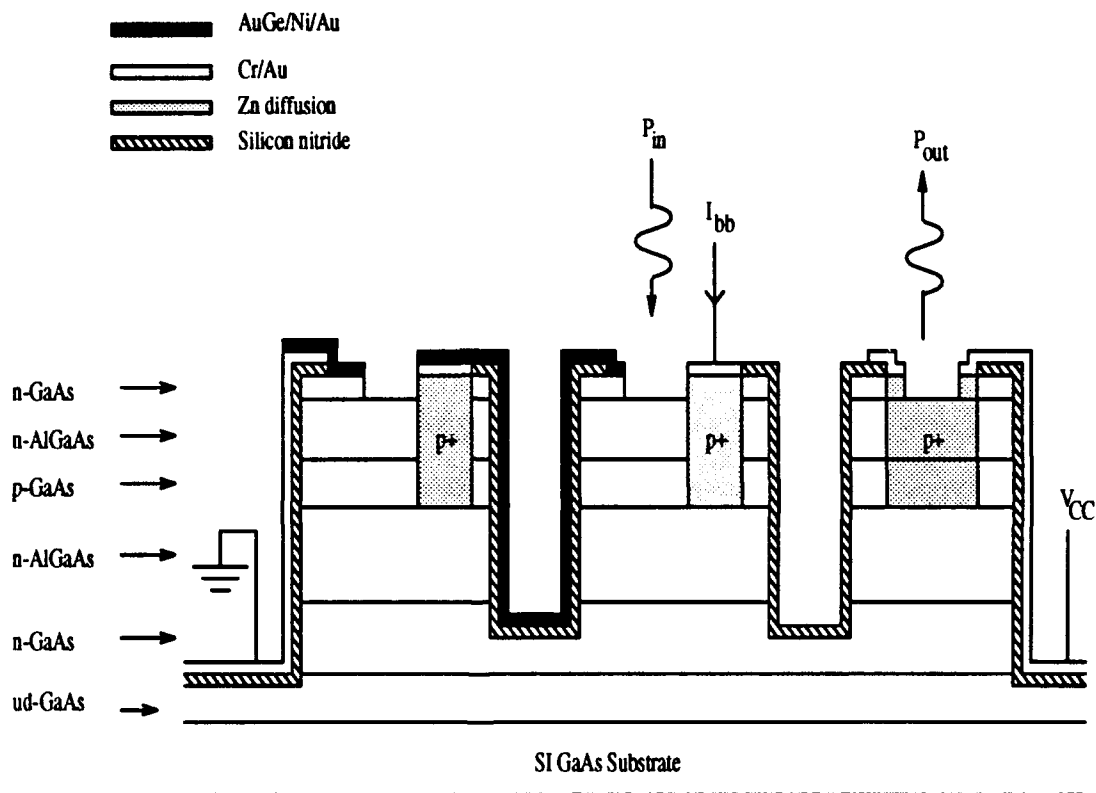
Fig. 2.7 Cross sectional view of the monolithically integrated optoelectronic neuron that is consisted of two Zn-diffused double-heterojunction bipolar transistors, which form a Darlington transistor pair, and a LED.

Epitaxial layers

n-GaAs

n-AlGaAs

p-GaAs

n-AlGaAs

n-GaAs

ud-GaAs

SI GaAs Substrate

Cell isolation

n-GaAs

n-AlGaAs

p-GaAs

n-AlGaAs

n-GaAs

ud-GaAs

SI GaAs Substrate

Device isolation

n-GaAs

n-AlGaAs

p-GaAs

n-AlGaAs

n-GaAs

ud-GaAs

SI GaAs Substrate

Fig. 2.8(a) Fabrication process of the optoelectronic neuron, illustrating the cell isolation and the device isolation.
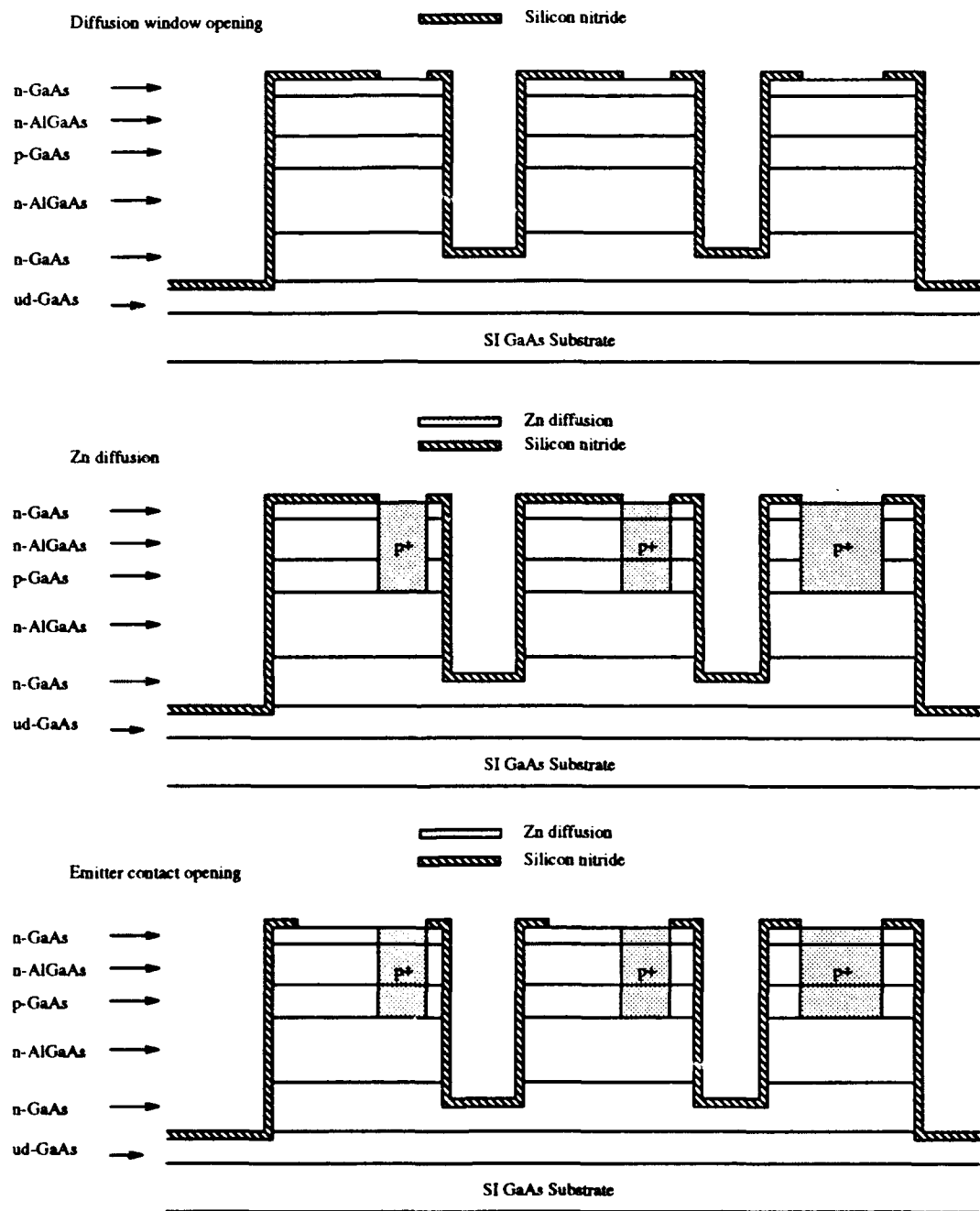
Fig. 2.8(b) Fabrication process of the optoelectronic neuron, illustrating the process of Zn-diffusion and emitter contact opening.

Diffusion window opening   ▨▨▨▨▨ Silicon nitride

n-GaAs   ⟶
n-AlGaAs   ⟶
p-GaAs   ⟶
n-AlGaAs   ⟶
n-GaAs   ⟶
ud-GaAs   ⟶

SI GaAs Substrate

                          ▭▭▭▭ Zn Diffusion
                          ▨▨▨▨ Silicon nitride
Zn diffusion

n-GaAs   ⟶
n-AlGaAs   ⟶
p-GaAs   ⟶
n-AlGaAs   ⟶
n-GaAs   ⟶
ud-GaAs   ⟶

SI GaAs Substrate

                          ▭▭▭▭ Zn Diffusion
                          ▨▨▨▨ Silicon nitride
Emitter contact opening

n-GaAs   ⟶
n-AlGaAs   ⟶
p-GaAs   ⟶
n-AlGaAs   ⟶
n-GaAs   ⟶
ud-GaAs   ⟶

SI GaAs Substrate
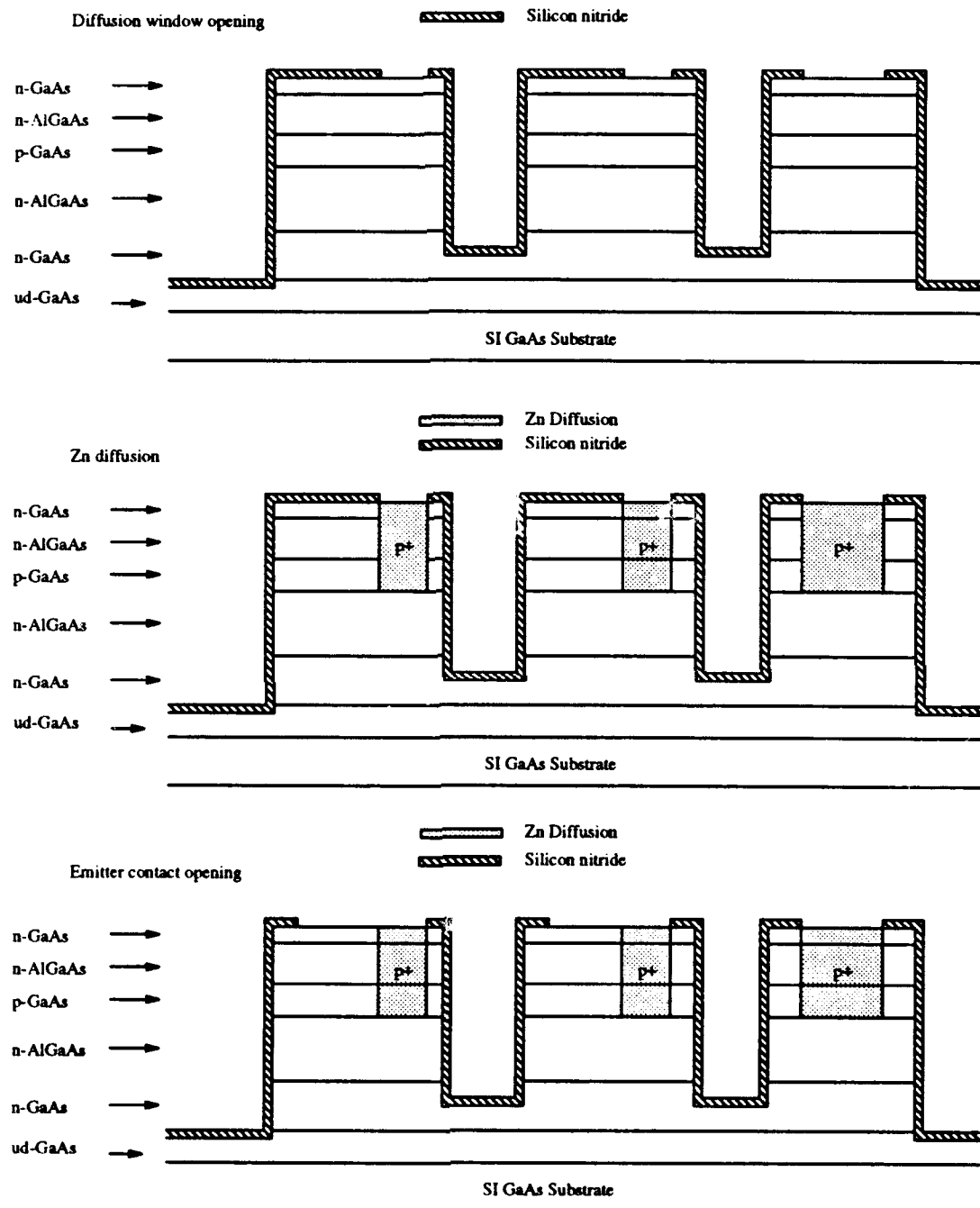
Fig. 2.8(c) Fabrication process of the optoelectronic neuron, illustrating the process of p-type and n-type metalizations.

some of the neurons were not electrically connected. This was purposely designed so as to avoid the "host" image which would have been created when the output of one neuron was diffracted by the grating intended for its neighboring neuron.

When the integrated optoelectronic neuron was tested, semiconductor controlled rectifier (SCR) characteristics were observed as shown in Fig. 2.9. A forward breakdown voltage of 75 V, a forward holding voltage of 25 V and a reverse breakdown voltage of 60 V were measured. By either increasing the base current or the external illumination on the neuron, the forward breakdown decreased. The LED was observed to emit light in the forward breakdown mode, implying the carriers were recombing in the low bandgap GaAs layer.

Careful inspection of the integrated LED with Darlington transistor pair revealed that the SCR, shown in Figure 2.10, was present in the device due to the parasitic p-n-p transistor coupled to the n-p-n DHBT. The anode of the SCR was the Zn-diffused area in the original LED region and the cathode was the original emitter (ground). This parasitic p-n-p transistor existed because the LED and the Darlington transistor pair shared the same collector. The effective base width of this parasitic p-n-p transistor was at least the separation between the LED and the bipolar transistor, which was 20 μm.

A reliable and efficient method would be to electrically isolate the LED from the Darlington transistor pair and then employ metalization to connect them as required. This method was pursued in our later version of the optical neurons. The process involved was to perform an additional isolation etch between the LED and the transistors down into the semi-insulting GaAs substrate, followed by evaporation of n-type metalization to appropriately connect the LED and the transistors up. The device cross sectional view after the remedial process was applied is shown in Fig. 2.11.

By etching into the semi-insulating substrate and employing metalization to
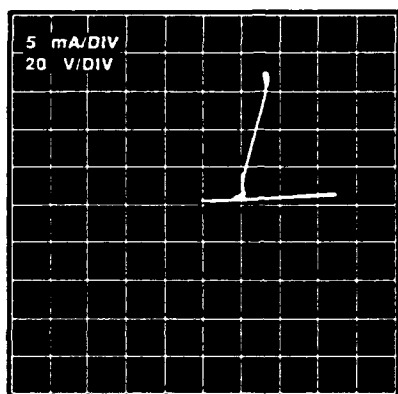
Fig. 2.9 I-V characteristics of the integrated optoelectronic neurons, exhibiting the behavior of a semiconductor controlled rectifier (SCR).
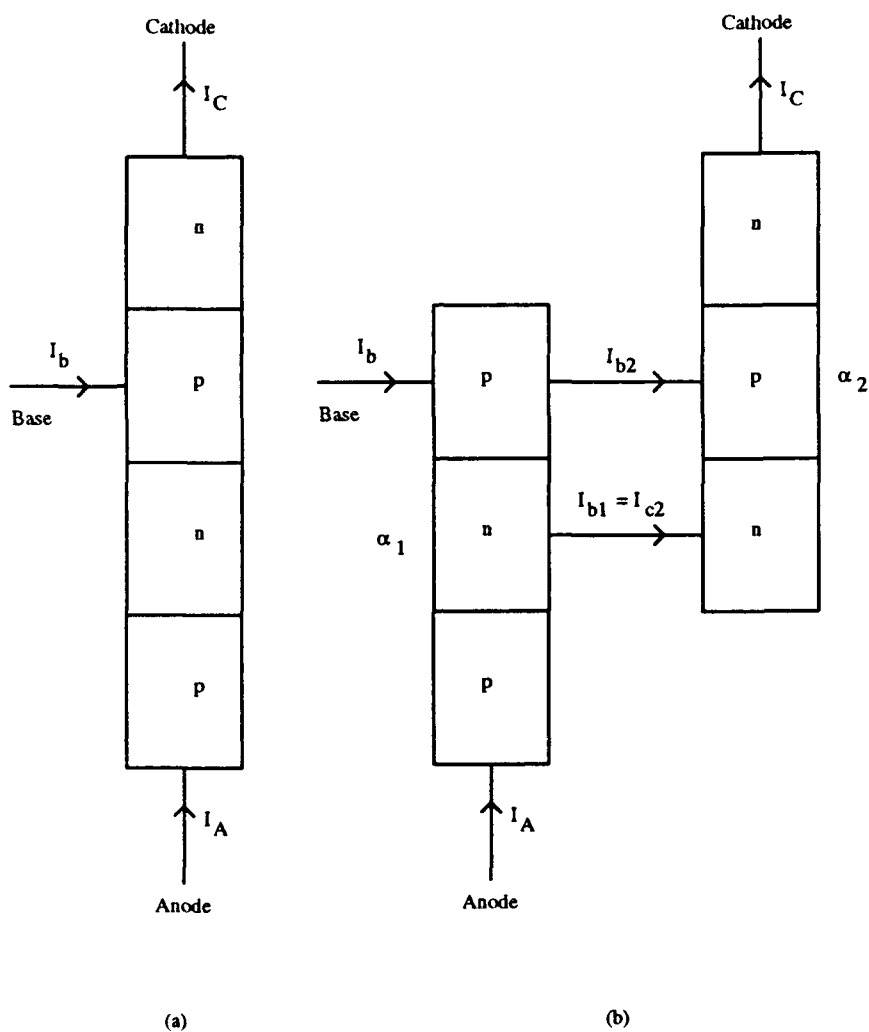
Fig. 2.10 (a) The structure of a semiconductor controlled rectifier, consisting of alternating p-n-p-n layers. (b) The device model of an SCR, illustrating that a SCR can be modeled as being composed of a p-n-p and n-p-n transistor connected in a fashion as shown in the figure.
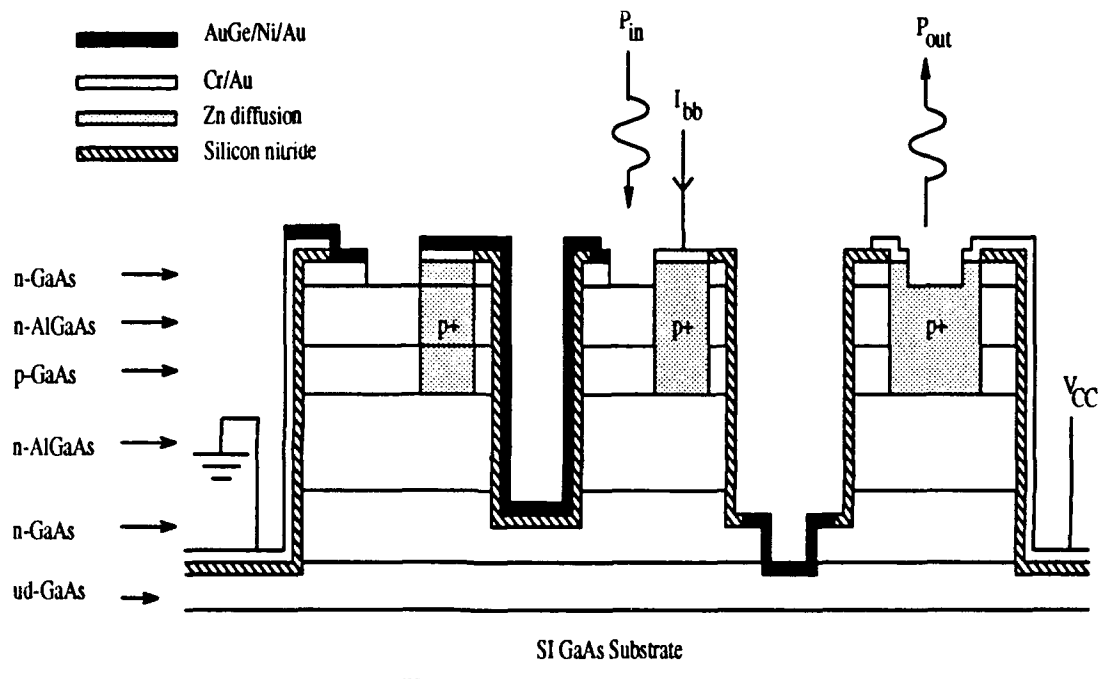
Fig. 2.11 The cross sectional view of the optoelectronic neuron after the the LED has been isolated from the transistors by an etch into the substrate and subsequently connected to the transistors by metalization.

connect the LED to the transistors, the optoelectronic neuron showed the correct I-V characteristics as shown in Fig. 2.12, which shows the common-emitter I-V characteristics for the Darlington transistor pair monolithically integrated to the LED. Even though the combined current gain was measured to be 2 at best, it exhibited the proper transistor characteristics offset by a voltage caused by the turn-on voltage of the LED. Thus, the origin of thyristor latching was successfully verified to be due to the parasitic coupling of the p-n-p transistor. This has an important consequence on the design of the integration. Namely, bipolar devices, such as bipolar transistors, LED's, and lasers, should be totally isolated from one another when integrating these devices on a common substrate. The integrated laser and bipolar transistor on an n^+ GaAs substrate as demonstrated by Katz et al. [10] might eventually be limited by the thyristor latching.

The low current gain measured from the Darlington transistor pair suggested that base leakage current dominated the current transport in the base, which led to inefficient electron injection from the emitter. As a result, most of the base current was recombining either through the surface or inside the depletion region. To prevent carriers from recombining in these regions, the method of etching down to the depleted AlGaAs in area between the base and the emitter of the transistor has been introduced and analyzed. Figure 2.13 shows the region of the Darlington transistor pair to be etched down. Since the region to be etched was the only region that was exposed to air, no extra mask was needed as the etching was done in a self-aligned manner. By applying this technique to the current monolithically integrated Darlington transistor pair, a plot of the combined current gain vs. the etch depth can be obtained. This is shown in Fig. 2.14.

It is worthwhile to note that, from Fig. 2.14, the dramatic improvement obtained in the current gain as the isolation etch depth increased was similar to the improvements obtained in individual discrete transistor and suggested the same
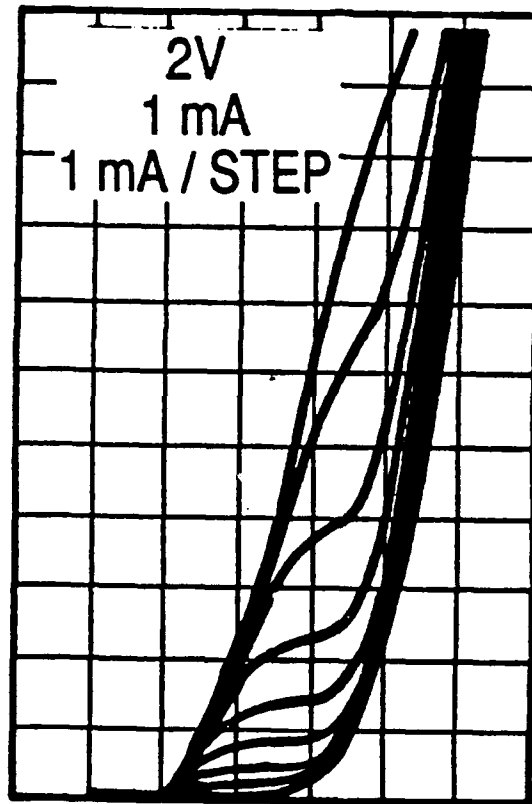
Fig. 2.12 I-V characteristics of the monolithically integrated Darlington transistors and LED after the LED has been electrically isolated from the transistors. The offset voltage in the V_{CE} of 2 V was due to the combination of the turn-on voltage of the LED and the intrinsic offset voltage in the transistors.
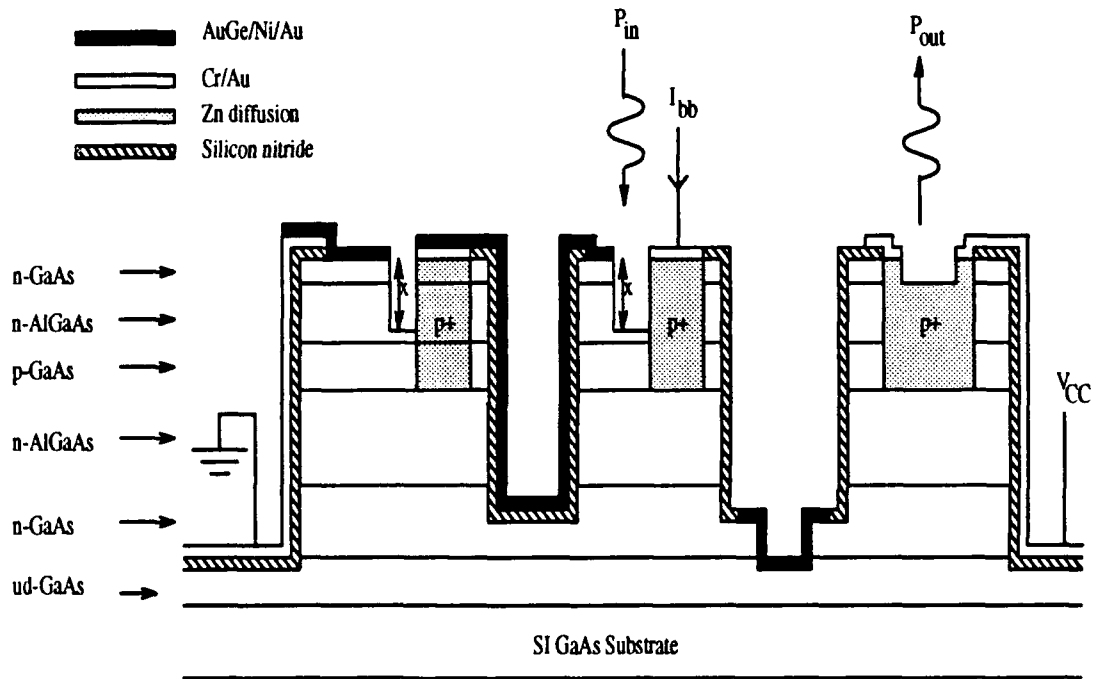
Fig. 2.13 Cross sectional view of the Darlington transistor pair monolithically integrated with the LED. In order to maximize the current gain of the transistors, the bulk emitter region between the emitter and the base contacts were etched away to leave only a thin and depleted AlGaAs as a passivation layer.
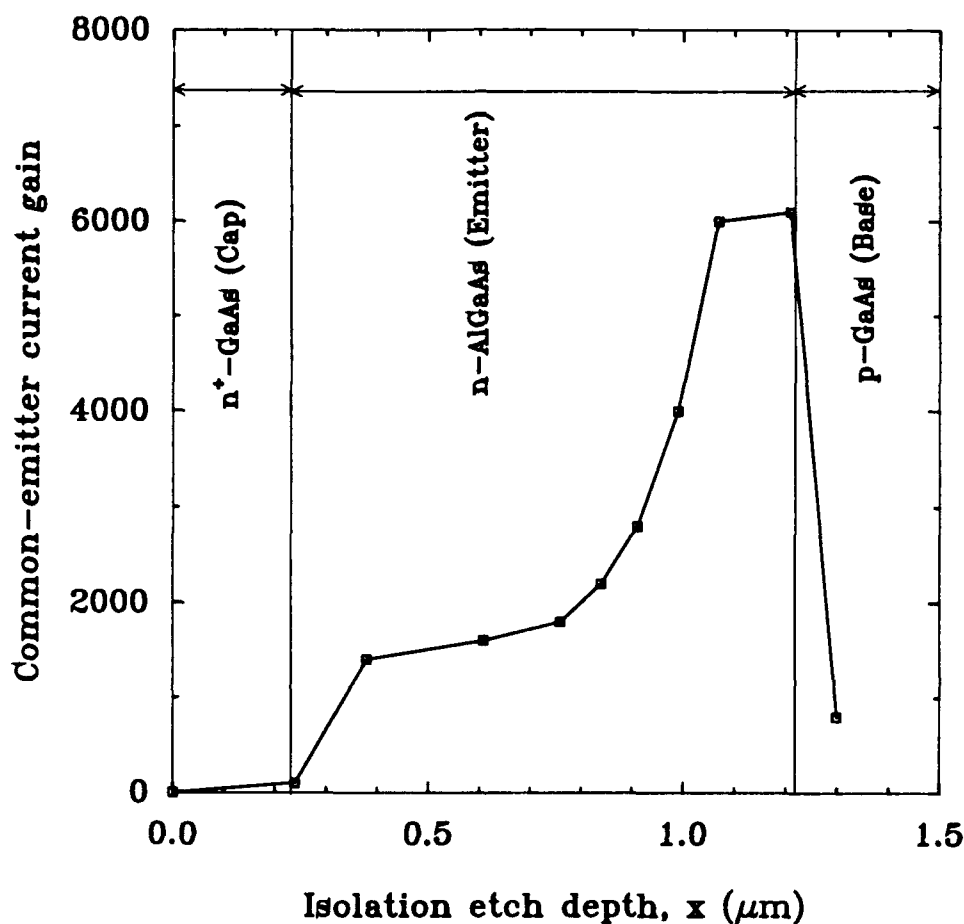
Fig. 2.14 Common-emitter current gain as a function of the isolation etch depth. x corresponds to the etch depth depicted in Fig. 2.13. Before the etch, the current gain was only 10, indicating the majority of the base current was flowing through the n^+ GaAs cap layer. As the etch depth increased, the current improved steadily until the remaining AlGaAs emitter was totally depleted. At which point, the current gain saturated. Thereafter, the current degraded dramatically due to the exposure of the base to the air, which promoted the surface recombination current.

mechanism by which the base current was transported in the transistor. Before the etch, the current gain was only measured to about 10. This was an indication that the majority of the base current was flowing through the n^+ GaAs cap layer. As the etch front penetrated into the AlGaAs emitter layer, the current gain improved steadily until the remaining AlGaAs layer was totally depleted, which, in turn, served as a surface passivation layer, and effectively eliminated all the recombination current that flowed on the surface. This was clearly illustrated by the flat plateau that indicated the saturation of the current gain at this situation. The width of the plateau approximately corresponded to the maximum thickness of the AlGaAs emitter layer that would be used to passivate the surface. As long as there was a depleted AlGaAs layer covering the base layer, the current remained constant, suggesting the independence of the current gain by the surface effect. However, as the isolation etch reached into the base layer, the extrinsic base region was exposed to air, thereby enhancing the recombination of electrons and holes on the exposed surface. This resulted in a detrimental reduction in the current gain of the transistor as the current gain plummeted to about 10% of the maximum value. By using this technique, a maximum current gain of 6000 was obtained in the Darlington transistor. This would more than satisfy the loop gain requirement imposed by the network. However, the current level at which this gain of 6000 was measured was at 20 mA. With a 5-volt power supply, the electrical power dissipation was 100 mW. Without a special cooling design, the heat generated would seriously limit the density of the neurons as eventually the generated heat would cause the device to fail. Thus, unless the same current gain could be obtained at however a much lower current level, the integration density for the neurons that were based on bipolar transistors would be severely hampered.

Metal Semiconductor Field-Effect Transistor-Based Neurons

3.1 Introduction

In the previous section, we described the integrated DHBT-based optoelectronic neurons. Because of the high electrical power dissipation by the neuron, it is concluded that a large array of these neurons would present a very severe heat dissipation problem, which eventually would lead to the failure of the chip. This limitation originates from the fact that the current gain of a bipolar transistor depends on how hard the transistor is driven. The larger the current is, the higher the current gain will be. This has the undesirable consequence of obtaining the gain required at the expense of power dissipation. In addition, this circuit does not have any input-output isolation. The input signal is directly amplified to obtain the output signal. Thus, any gain required will have to be directly provided by the Darlington transistor pair. This puts a very stringent requirement on the Darlington transistor pair. Namely, it has to provide a sufficient current gain and yet dissipate little power. This however contradicts the gain-power tradeoff rule stated earlier. Thus, an alternative design that decouples the relationship between the gain and the power as well as the relationship between the input and the output has to be developed.

One such possible alternative is to use a voltage-controlled transistor, such as a metal-semiconductor field-effect transistor (MESFET), to drive the LED. This MESFET is in turn controlled by an input switching circuit, composed of a detector, which accepts the input light, and a loading transistor. Thus, as the detector detects a sufficient input light, it pulls up the loading transistor. As a result, the LED-driving MESFET is turned on and it drives the LED. The advantage of this
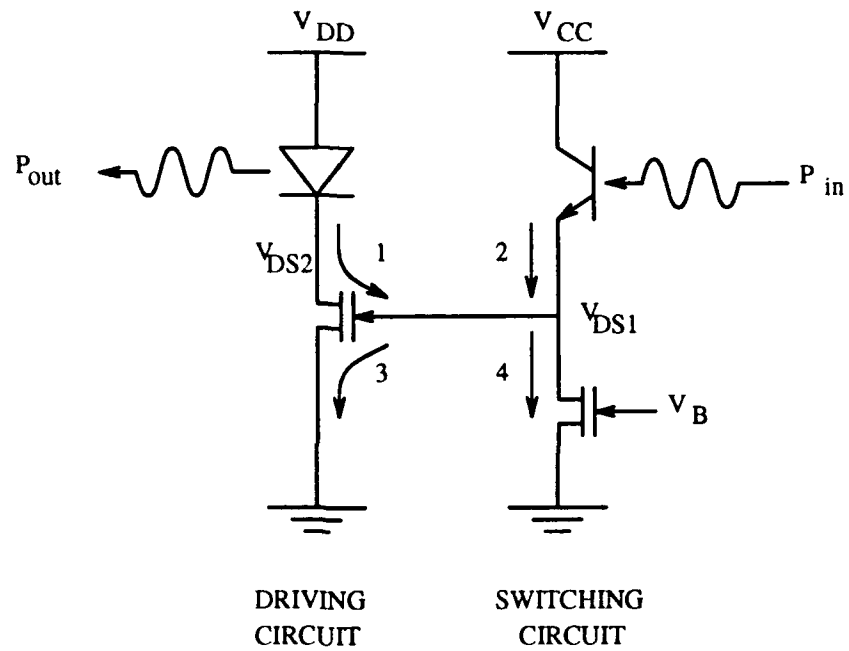
Fig. 3.1 Schematic circuit diagram of the optoelectronic neuron that incorporates two MESFET's, a phototransistor and a LED.

circuit design is that the LED is indirectly controlled by the input light, as opposed to the direct amplification of the input to produce the output in the DHBT-based neuron. Figure 3.1 depicts one such possible circuit, where a phototransistor is used as the detector and another MESFET is used as the loading transistor. As seen in Fig. 3.1, this circuit can be divided into the output driving circuit and the input switching circuit. The gate of the output driving MESFET is controlled by the voltage between the phototransistor and the loading MESFET. This voltage fluctuates between ground and V_{CC}, depending upon the the photocurrent generated. As the input optical power increases, the photocurrent increases. At a certain point, the generated photocurrent has surmounted the current drawn by the loading MESFET. At this point, this voltage changes from ground to V_{CC}. This turns on the output driving MESFET. A direct consequence of this circuit design is that the isolation of the output and the input. This would enhance the sensitivity of the circuit as the circuit can be designed to switch by a very weak input light. Another consequence of this switching action is that the optical gain is now determined by the relative output impedance of the phototransistor and the loading MESFET. As will be shown later, if these two transistors have infinite output impedances, the neuron can be turned on instantaneously. Other advantages of this circuit include the relatively mature technology in fabricating the high-gain MESFET's and a much lower electrical power dissipation required to turn on the neuron.

3.2 Analysis of MESFET-Based Neurons

The MESFET-based optoelectronic neuron consists of an output driving circuit that is relatively simple in circuit design and easy to understand. The LED is driven by a MESFET, which, in turn, is controlled by an input switching circuit. To fully appreciate this circuit, one has to understand how the input switching
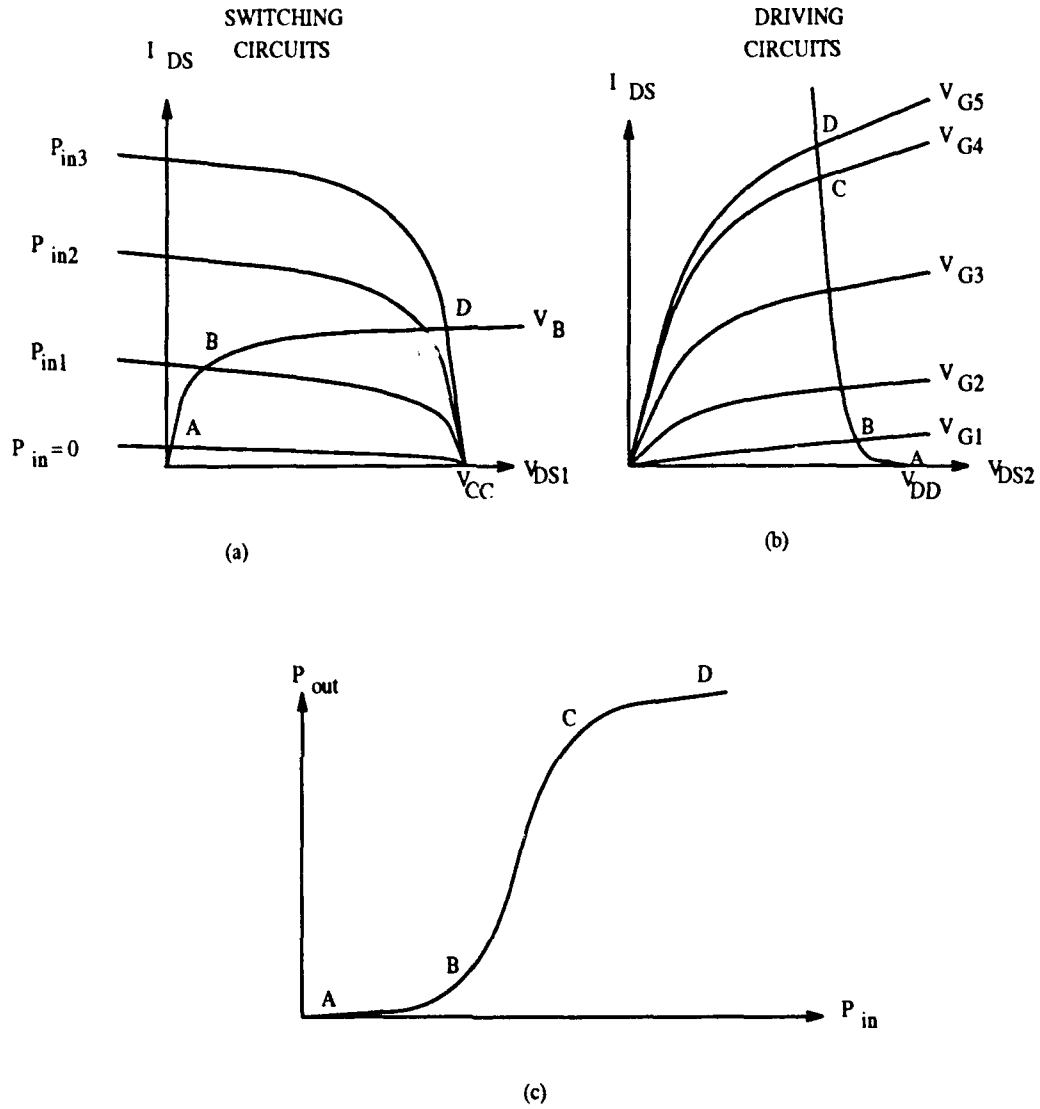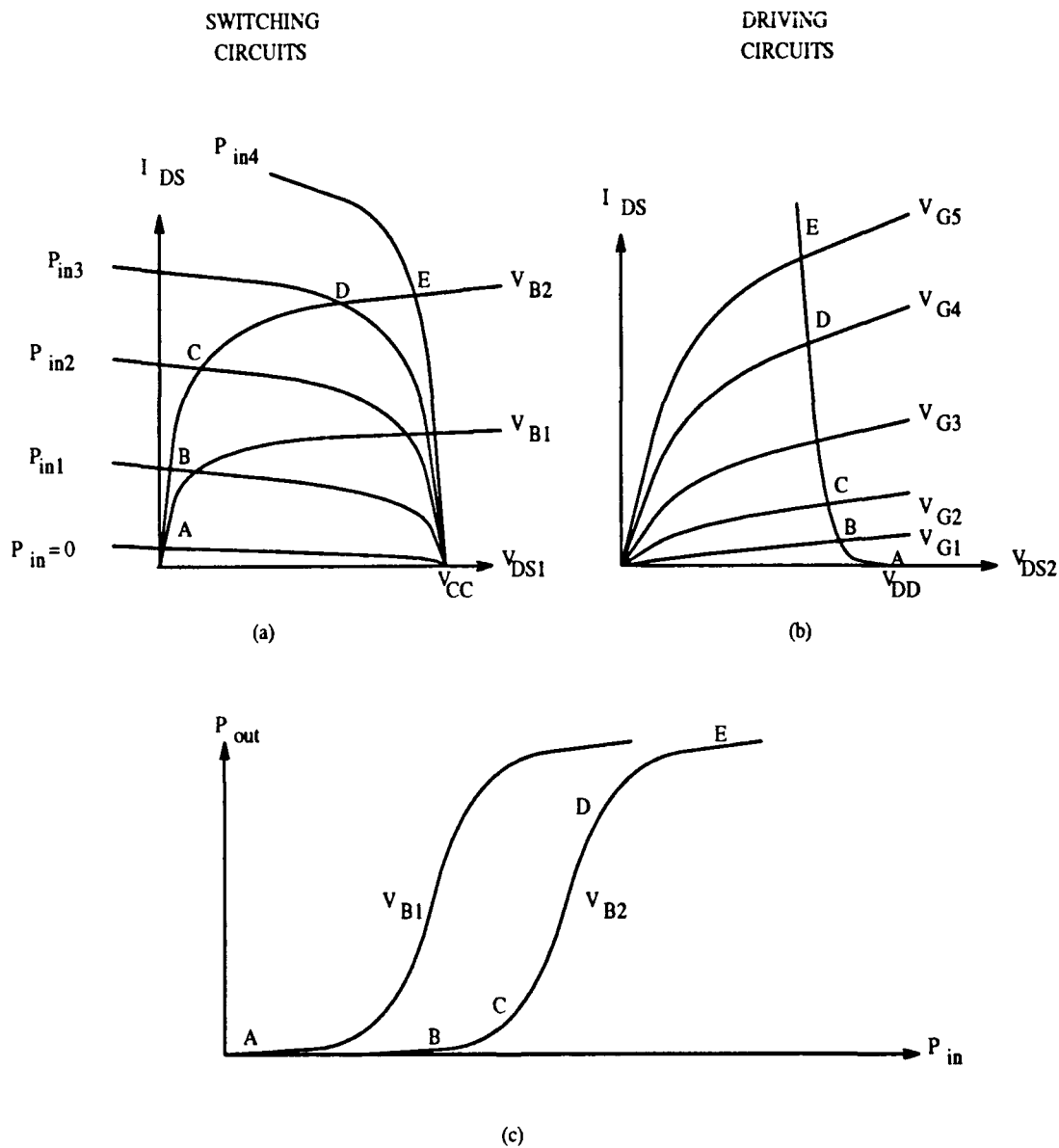
Fig. 3.2 (a) I-V characteristics of the input switching circuit in the MESFET-based neurons. The intersection point determines the operating point of the neuron. (b) I-V characteristics of the output driving circuit, which determines the output power level emitted by the LED. (c) Overall input-output characteristics of the neuron.

circuit works and how it affects the operation of the output driving circuit. If we restrict ourselves first to analyze the input switching circuit, we see that the voltage in the middle of the two devices, V_{DS1}, can be designed to swing between ground and V_{CC}, depending on the relative currents drawn by each device, namely the phototransistors and the loading MESFET. This is best understood by analyzing the I-V characteristics of both transistors plotted on the same graph. Shown in Fig. 3.2(a) is one such plot. The I-V curve for the loading MESFET is plotted in a conventional way with the vertical axis being the drain-source current and the horizontal axis being the drain-source voltage, which is labeled as V_{DS1}. In order to plot the I-V curve for the phototransistor on the same plot, we use the fact that the voltage across the emitter and the collector, V_{CE}, is given by $V_{CC} - V_{DS1}$. Therefore, the I-V curve for the phototransistor is first flipped with respect to the vertical axis to get the $-V_{DS1}$ and then linearly translated to the right by V_{CC}. The resultant plot is shown Fig. 3.2(a). The voltage at the middle node, V_{DS1}, which is also the gate voltage of the output MESFET, is determined by the intersection point of the two transistor curves. For input light power equal to zero, the value of V_{DS1} is almost equal to zero, as indicated by point A in the figure. However, as the input light power gradually increases from zero to P_{in3}, the voltage, V_{DS1}, changes from point A through point B and point C and to point D. Thus, V_{DS1}, swings from almost ground to almost V_{CC}. To see how this swing in V_{DS1} affects the output circuit, a similar I-V curve of the output driving MESFET with the LED plotted backward is shown in Fig. 3.2(b). The swing in V_{DS1} corresponds to a swing in the gate voltage, which is designated by V_{G1} through V_{G5}. Assuming the output driving MESFET is an enhancement-mode transistor, the initial V_{DS1}, or the gate voltage, of zero volt (point A) will not turn on the transistor. Thus the transistor is still in cut-off as indicated by point A in the output I-V curve. Since there is no current between the source and the drain, the LED is off. As V_{DS1} swings from ground to

the voltage corresponding to point D, the gate voltage changes from ground to V_{G5}, which puts the MESFET in a strong forward conduction mode. The current flowing between the source and the drain is used to drive the LED, which emits light with an intensity that is linearly proportional to the current which passes through the LED. If we take the output power emitted by the LED as the output and retain the input power to the phototransistor as the input, we obtain the input-output characteristics of the MESFET-based optoelectronic neuron. This is shown in Fig. 2(c). Points A, B, C, and D are also labeled to indicate the various states that the neuron is in. It should be noted that the output power from the LED does not increase too much as the input power increases from 0 to P_{in1} (from point A to point B). This is because the increase in the input power does not generate enough photocurrent in the phototransistor to cause a significant change in V_{DS1}. However, from point B to point C, a dramatic increase in the LED output power is observed. This is due to the large change in V_{DS1} which is, in turn, caused by the phototransistor current overtaking the current drawn by the loading MESFET. This dramatic increase in the LED output power simulates the thresholding characteristics in the neurons with the level of threshold controlled by the biasing voltage, V_B, which is the gate voltage of the loading MESFET in the input switching circuit. From point C to point D, there is only a small change in V_{DS1}. Thus, the change in the LED output power is small. This provides a saturation effect, which is desirable for simulating the thresholding operation of the neurons. Therefore, the thresholding and saturation behaviors of the neurons can be easily controlled and simulated by using these four devices.

The level of the threshold can be adjusted by applying a different voltage to the gate of the loading MESFET in the input switching circuit. The various operating states labeled by A through E are illustrated in Fig. 3.3 (a)-(c). From these plots, it is clear that this circuit does not only provide the desired input-output

Fig. 3.3 (a) I-V characteristics of the input switching circuit in the MESFET-based neurons with a different biasing voltage applied to the gate of the loading MESFET. (b) I-V characteristics of the output driving circuit, which determines the output power level emitted by the LED. (c) Overall input-output characteristics of the neuron.

characteristics for the neurons, but it also enables the threshold level of the neuron to be electronically tuned through this bias terminal. This feature will be needed for the dynamics in a massively interconnected network of neurons. Generally speaking, the higher the biasing voltage, V_B, the higher the level of threshold in the neuron since a higher V_B induces a higher source-drain current through the MESFET, which, in turn, requires a higher input power in order to establish the onset of the threshold. From the physical standpoint, this loading MESFET provides a reference current against which the generated photocurrent from the phototransistor is compared. If the photocurrent generated by the phototransistor is not sufficient to meet the reference current, the voltage at V_{DS1} is pinned to ground. However, if the photocurrent is greater than the reference current, the voltage at V_{DS1} is pulled up to V_{CC}, which causes the switching to occur. By varying the magnitude of this reference current through V_B, one can obtain a set of thresholding input-output curves with different threshold levels.

The optical gain in the MESFET-based neuron is determined by the ability of the input circuit to switch for a given input intensity. Qualitatively, as long as the photo-generated current is larger than the reference current drawn by the loading MESFET, the switching occurs. However, there is a region in which the output rises gradually from zero to maximum. The slope of the rise defines the differentially optical gain. In this region, the output power level depends critically on the gate voltage of the output driving MESFET. If the gate voltage rises sluggishly, the output of the LED is expected to rise sluggishly as well. On the other hand, if this gate voltage rises instantaneously, the output of the LED rises instantaneously. Thus, the differential optical gain is determined by the sensitivity with which the voltage, V_{DS1}, can be raised for a given input light. This is further determined by the relative flatness in the I-V curves of both the phototransistor and the loading MESFET. If the output saturation currents of both transistors are not constant
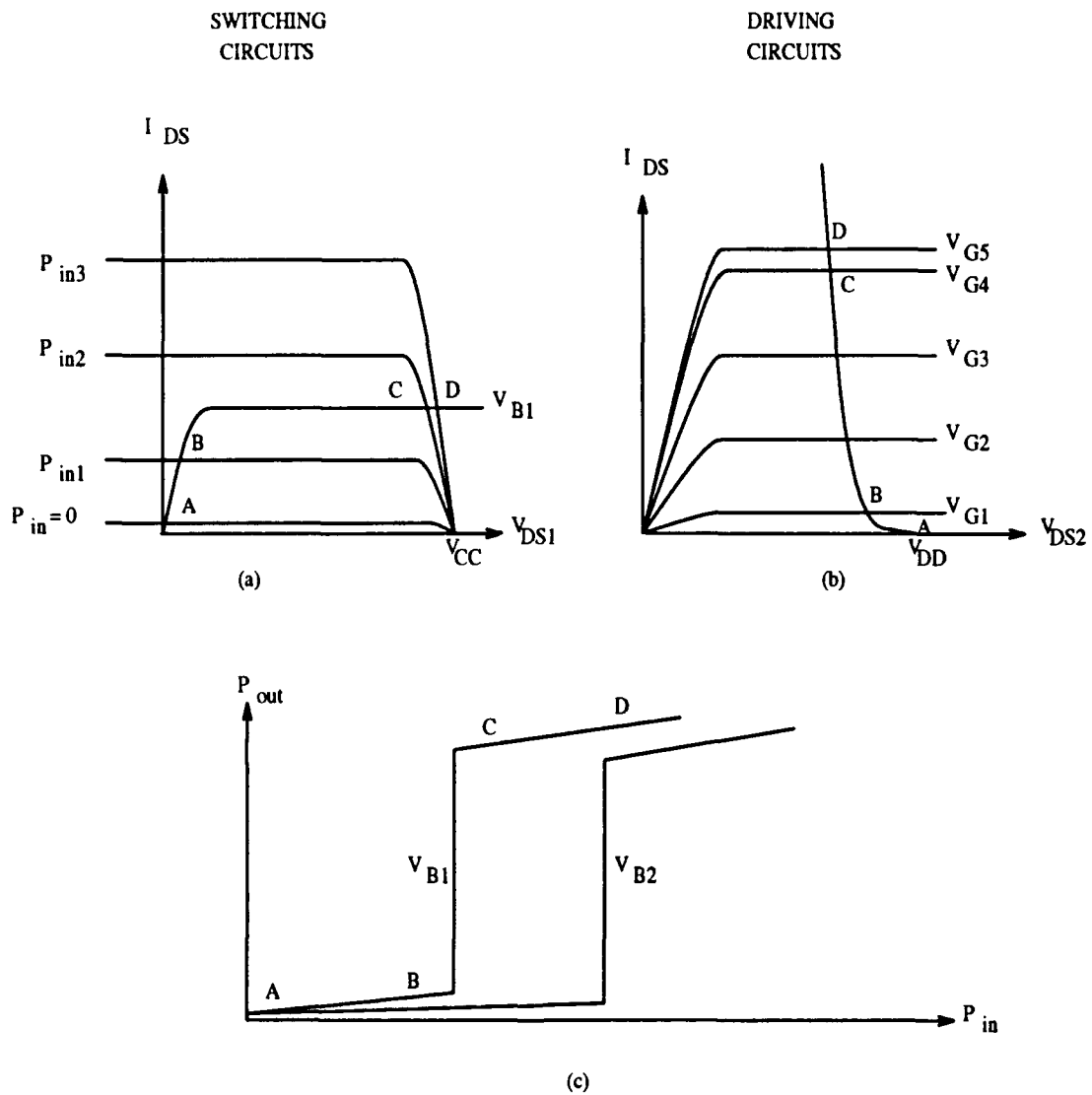
Fig. 3.4 (a) I-V characteristics of the input switching circuit in the MESFET-based neurons with an infinite output impedance in both the phototransistor and the loading MESFET. (b) I-V characteristics of the output driving circuit again with an infinite output impedance in the driving MESFET. (c) Overall input-output characteristics of the neuron showing an instantaneous switching as a result of having an infinite output impedance in the transistors.

as the voltage varies, the switching will be a soft one. This is because the rise in current in one transistor accompanied by the same rise in current in the other transistor has to be accomplished by a change in the transistor voltage. Otherwise, current continuity will not be satisfied. This is the case that corresponds to finite output impedances in the transistors. If, however, the output saturation currents of both transistors remain constant, the switching characteristics can be expected to be an abrupt one because the photo-generated current and the reference are now independent of the voltage across the transistors and a comparison of the relative magnitude of the currents will uniquely determine the state of the switch. If the photocurrent is slightly less than the reference current, the circuit will not switch. If, however, the photocurrent is just slightly larger than the reference, the circuit will switch. This instantaneous switching characteristic, caused by the infinite output impedances in the transistors, translates into an infinite differential optical gain. Therefore, it is extremely desirable to make these transistors with very high output impedances so that high-gain neurons can be obtained. The switching characteristics of the circuits for the infinite impedance case are illustrated in Fig. 3.4 (a)-(c) again with points A through D again to show the various states the circuit is in.

The nature of the output impedance can be quite complicated. It can be due to improper design in the material that causes the non-saturating current. For example, a low base doping concentration in the phototransistor will cause a severe sloping in the output current. However, increasing the base doping concentration unfortunately decreases the current gain, β, of the transistor. For MESFET's, the non-saturating output current is usually due to the source-drain current that spills into the substrate [20] and causes a bias-dependent source-drain current. The origin of the non-saturating output current can be also caused by the leakage current in the transistors, particularly the reverse leakage current across the gate and the drain.

As the voltage, V_{DS1}, is being raised from ground to V_{CC}, the gate-drain Schottky diode experiences a stronger reverse bias as the gate voltage is kept constant. This introduces a larger leakage current, which flows from the drain to the gate. From the point of view of V_{DS1}, this leakage current is no different from the reference current drawn by the same MESFET because both of these currents flow out of the node at V_{DS1}. As a result, this leakage current is mixed into the reference current, which is usually bias-dependent to start with. Therefore, the total current becomes even more bias-dependent and consequently the output impedance decreases.

Leakage currents complicate the analysis of the switching behavior significantly if the input switching circuit is connected to the output driving circuit. This is because the isolation between the input and the output circuits is not complete. Having a Schottky diode at the gate, MESFET's inevitably draw leakage current across the gate, either from the source or from the drain, depending on the bias of the transistor. The switching characteristics presented above is an overly simplified picture of the real device. In reality, there are 4 basic current components that determine the switching characteristics of the neuron (instead of just two as previously mentioned). Referring to Fig. 3.5(a), I_2 is the photo-generated current from the phototransistor, and I_4 is the reference current drawn by the loading MESFET. In addition, there is a current, I_1, that represents the leakage current across the gate and the drain in the output driving MESFET and an I_3 that represents the other leakage current component in the MESFET, which is the gate-source leakage current. At any time, the sum of I_1 and I_2 has to equal the sum of I_3 and I_4 in order to satisfy the current continuity equation. Since these four current components depend on V_{DS1}, V_{DS1} will adjust itself such that the current continuity equation is satisfied. Any perturbation in any one of the current components will cause the re-adjustment in V_{DS1}. As the input power is increased, I_2 increases proportionally. Thus, V_{DS1} reacts to this imbalance by increasing its value, which, in turn,
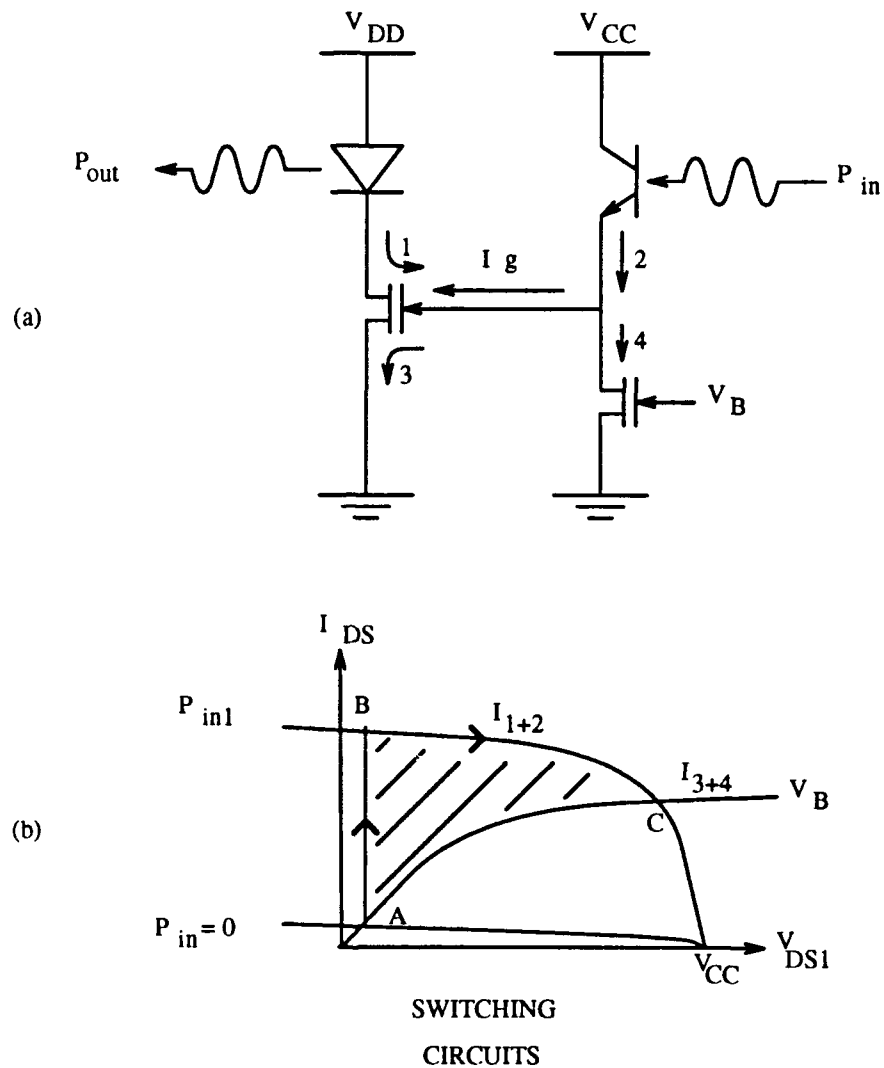
Fig. 3.5 (a) Schematic circuit diagram of the MESFET-based neuron showing the four current components that affect the gate voltage of the output driving MESFET, V_{DS1}. (b) Evolution of current dynamics as the neuron is being turned on. The neuron starts from point A in the off-state and ends at point C in the on-state through the state at point B. The shaded region represents the excess current available to charge the gate of the output driving MESFET.

decreases I_2 and I_1 and increases I_3 and I_4 at the same time so that the current continuity is satisfied again. As far as V_{DS1} is concerned, there is no difference between I_1 and I_2 because these two current components both flow into the node, providing the excess carriers needed by the other two current components. Nor is there any difference between I_3 and I_4 from the standpoint of V_{DS1} because these two current components both flow out of the same node, removing carriers that are injected by I_1 and I_2. Thus, at the end, we can still treat this switching circuit as being consisted of two current components; one flowing into the gate of the output driving MESFET and the other one flowing out of the gate. When the input light illuminates on the phototransistor, there will be an excess current that flows in the gate. This excess current is used to charge up the capacitance associated with increasing the gate voltage to its proper value. However, as the gate voltage increases, the magnitude of the excess current decreases owing to a smaller current that flows into the gate and a larger current that flows out of the gate. Eventually, as the final gate voltage is established, the current flowing into the gate is again equal to that flowing out of the gate. This process is illustrated in Fig. 3.5(b). Initially, the neuron is in the off-state, which is indicated by point A. As the input power jumps from zero to P_{in1}, the current that flows into the gate, $I_1 + I_2$, all of sudden increases to a value dictated by the amount of the input power, labeled as point B. This increase can not be accommodated immediately by the current that flows out the gate, $I_3 + I_4$. Therefore, the gate voltage has to increase in trying to balance the two current components. However, the gate voltage can not be raised immediately because these is a capacitance associated with charging up the gate. As a result, this excess current goes to charge up the capacitance of the gate in bringing up the gate voltage until the the two current components balance each other. The time over which this switching takes place depends on the relative magnitude of the two current components. From this plot, it can be easily inferred that the stronger the

input power is, the faster it will be for the neuron to reach the steady-state because there is more excess current available to charge up the gate. In fact, the switching time can be found by solving for τ_{charge} in the following equation.

$$CV_{CC} \approx \int_0^{\tau_{charge}} (I_1 + I_2 - I_3 - I_4)dt, \qquad (3.1)$$

where C is the total gate capacitance that needs to be charged up. From this equation, we see that in order to decrease the switching time, one needs to decrease the capacitance and V_{CC} and increase the input power.

With this circuit, it is sometimes possible to have a situation in which the neuron is already on without any input power. This is due to the fact that I_1 is so large that it overcomes the combined currents of I_3 and I_4. As a result, the gate is fully charged up to almost V_{CC} and the LED is emitting. This situation is especially likely to occur when the output driving MESFET is very wide and the input loading MESFET is very narrow. The narrow-width MESFET is needed to increase the sensitivity of the input circuit. Thus, there is an optimal width in the loading MESFET that will prevent this phenomenon from happening and yet provide sufficient sensitivity. When this problem is present, it can be cured by increasing the biasing voltage, V_B, applied to the gate of the loading MESFET. This will increase the reference current, which provides a sink for I_1 to bring V_{DS1} down to the ground in order to shut the neuron off. In an opposite situation where the neuron can not be turned on by the input power, a bias optical beam can be applied to the phototransistor to generate more photocurrent, I_2. The magnitude of this optical beam can be just sufficient to bias the neuron to a point that the original input power will be able to turn on the neuron. This situation is illustrated in Fig. 3.6. Thus, by using either the electrical bias to increase the reference current so that the neuron can not be turned on without any input power or the optical bias

to decrease the amount of input power needed to turn on the neuron, the MESFET-based neuron can be properly tuned for maximum sensitivity and fault-tolerance.

3.3 Characterization of Discrete Devices

3.3.1 Metal Semiconductor Field-Effect Transistors

Metal semiconductor field-effect transistors are three-terminal devices in which one of the terminals, the gate, is used to control the current flow between the other two terminals, the source and the drain. The operational principle of the MESFET's is very similar to that of the junction field-effect transistors (JFET's) [21] except in MESFET's, Schottky diodes, as opposed to p-n diodes, are used to control the width of the depletion region, beneath which the current flows. In addition, because of the nature of the Schottky diodes, leakage currents through the gate tend to be higher in MESFET's as compared to those in JFET's. Nevertheless, the fabrication of MESFET's is much simpler because the formation of the control terminal, gate, is by metalization rather than by diffusion as in the case for JFET's.

A typical MESFET has one of the structures shown in Fig. 3.7. The first structure, shown in Fig. 3.7(a), is the simplest. It basically involves metalizing the source, the drain, and the gate appropriately on a properly doped material. The drawback is the relatively low breakdown voltage between the gate and the drain. Another disadvantage of this structure is the difficulty in placing the gate down accurately between the source and the drain. Since the spacing between the source and the drain is typically less than $10\mu m$ and the gate length is already few μm long, a tight control on photolithography is very crucial. As a result, a self-aligned structure, such as the one shown in Fig. 3.7(b), has been developed [22,23].
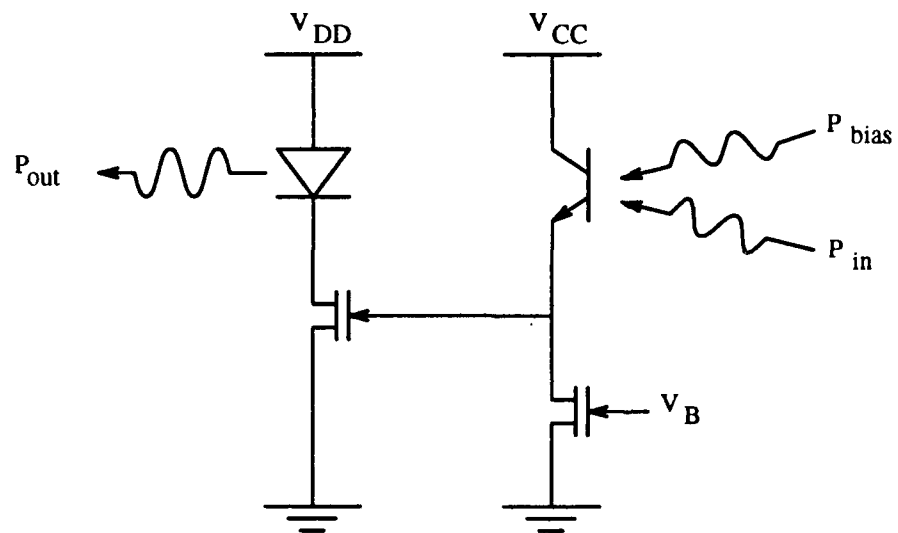
Fig. 3.6 Sensitivity of the MESFET-based neuron can be increased by applying an external optical beam to bias the neuron to just right before the threshold.
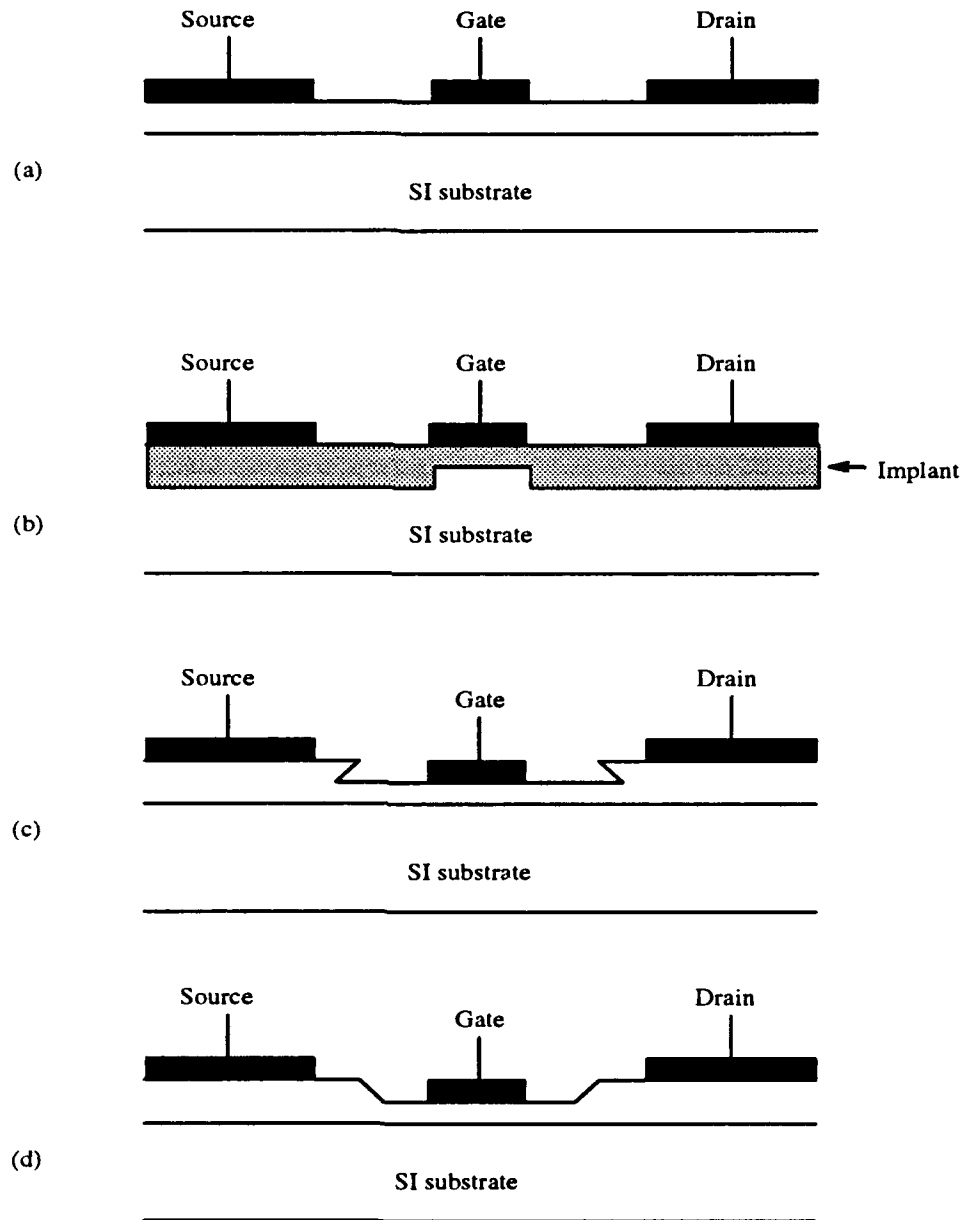
Fig. 3.7 Typical structure of a MESFET :(a) Conventional structure with planar geometry. (b) Self-aligned implanted structure. (c) & (d) Recessed gate structure.

It involves implanting an appropriate dosage of n-dopants into the material first to define the thickness and the doping level of the MESFET conduction channel layer. Then a special refractory gate material, typically made of Ti/Pt/Au, is evaporated, followed by a deeper implantation with a stronger dosage to define the ohmic regions for the drain and the source. This step is accomplished in a self-aligned manner because the gate metals are used as the implantation mask. Once the highly conductive ohmic regions for the source and the drain are defined, the actual source and drain metalizations can be defined without much precise control as long as they fall within the implanted region. Though the process is more tolerant, it does suffer from low breakdown voltage between the gate and the drain as the highly conductive ohmic drain region is very close to the gate. This is the direct result of using self-aligned implantation. However, if one designs the circuit properly so that the MESFET's will never be driven close to that breakdown voltage, this structure might prove to be very useful and yield very consistent device performance. In fact, this is the structure employed by the commercial MESFET company, such as Vitesse Semiconductor Corp [24]. Another way of making the MESFET, which will have a higher breakdown voltage, is to recess the gate slightly into the MESFET conduction channel layer by etching, such as the one shown in Fig. 3.7(c) and 3.7(d). Because of the property of GaAs, the side of the recess will make an obtuse or an acute angle with the surface depending on the orientation of the GaAs [25]. The effect of the recessed gate is not only to increase the breakdown voltage, but also to increase the transconductance of the MESFET through the reduction of the parasitic source-gate resistance [26]. However, between the structure in Fig. 3.7(c) and 3.7(d), the one in Fig. 3.7(c) tends to be less reliable as the sharp corners resulted from etching generate high electric fields locally around the corners. Thus, in this work, the MESFET structure in Fig. 3.7(d) is used.

For a recessed-gate MESFET, it is extremely crucial that the etch depth be

controlled as precisely as possible because the remaining channel will directly determine the pinch-off condition of the MESFET. Thus the operational mode of the MESFET, for example, either enhancement-mode or depletion-mode, will be affected by the amount of the recessed etch. For a recess that is shallow, the channel is not totally depleted. Therefore, a negative voltage is needed at the gate to pinch the channel off. This is the depletion-mode operation. On the other hand, if the recess is excessive such that the channel is already totally depleted, then a positive voltage is needed at the gate to induce a current flow between the source and drain. This is the depletion-mode operation of the MESFET.

Once the configuration of the MESFET is determined, there remains several detailed issues that need to be addressed for the optimization of the MESFET performance. Firstly, the parasitic resistance between the source and the gate contributes to a reduction in the effective voltage between the source and the gate. This can be explained by referring to Fig. 3.8. Because of the finite separation between the gate and the source, there is a parasitic resistance, R_{gs}, which accounts not only for this separation, but also the contact resistance and the distributed bulk resistance contributed by the source metalization. This resistance causes a voltage drop along the channel even before the channel current gets to the edge of the gate metalization. As a result, the effective voltage between the gate and the source is less. Therefore, the overall transconductance drops. The magnitude of the drop can be determined by the following expression.

$$
\begin{aligned}
g_m &= \frac{I_{ds}}{V_{gs}} \\
&= \frac{I_{ds}}{V'_{gs} + I_{ds}R_{gs}} \\
&= \frac{g'_m}{1 + g'_m R_{gs}},
\end{aligned}
\tag{3.2}
$$

Semi-insulating GaAs substrate
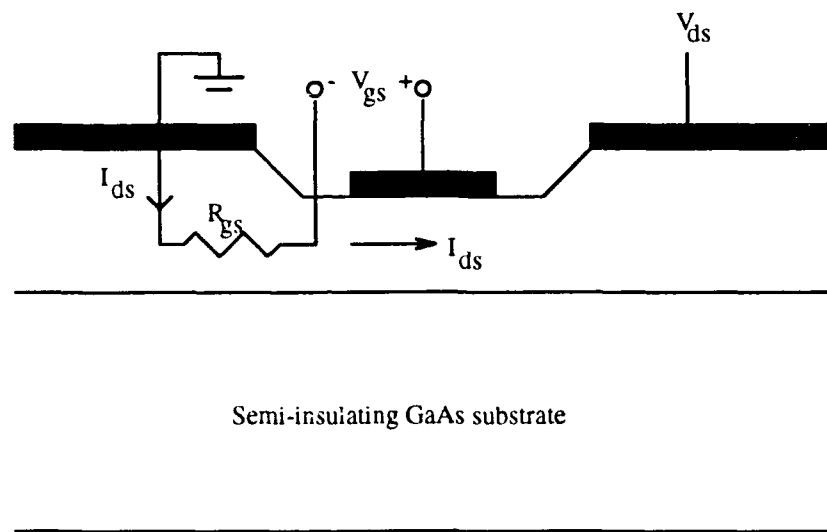
Fig. 3.8 Model of a MESFET including the parasitic R_{gs}, source-gate resistance.

where V'_{gs} and g'_m represent the true source-gate voltage and the intrinsic transconductance of the MESFET respectively. Therefore, the larger the parasitic resistance, R_{gs}, is, the more reduction there is in the transconductance of the transistor. This is very undesirable for the transistor. Thus, one should minimize this parasitic resistance. One way is to abridge the gap between the gate and the source metalizations. In the extreme case where the gap is zero, a self-aligned structure is obtained in which the edge of the gate metalization is aligned to the edge of the source metalization. This is illustrated in Fig. 3.9(a). This requires using the evaporated source metals as the etching mask in recessing the gate down to the appropriate depth. The area of the gate metalization has to also overlap slightly over the source metalization. This is because only by overlapping the two metalizations can a truly self-aligned structure be obtained. The overlapping portion of the gate metalization becomes part of the source contact with remaining non-overlapping gate physically defining the size of the gate metalization.

If we simply increase the length of the gate to overlap the source in obtaining the self-aligned structure, two problems arise. One is the degradation of the transconductance due to the increased gate length. The other one is the small breakdown voltage between the gate and the drain because of the small separation between the two terminals. In fact, the breakdown in a MESFET is usually dominated by the breakdown between the gate and the drain. If one measures the breakdown voltage of a MESFET, one would find it almost equal to the reverse breakdown voltage in the gate-drain Schottky diode. This is experimentally verified and shown in Fig. 3.10(a) and (b), in which the breakdown voltage of approximately 4 Volts in a MESFET shown in Fig. 3.10(a) matches well the breakdown voltage of its gate-drain Schottky diode shown in Fig. 3.10(b). This is indicative of the strong correlation between the two breakdown phenomena. To eliminate this problem, one has to place the gate farther away from the drain. Therefore, reducing the size
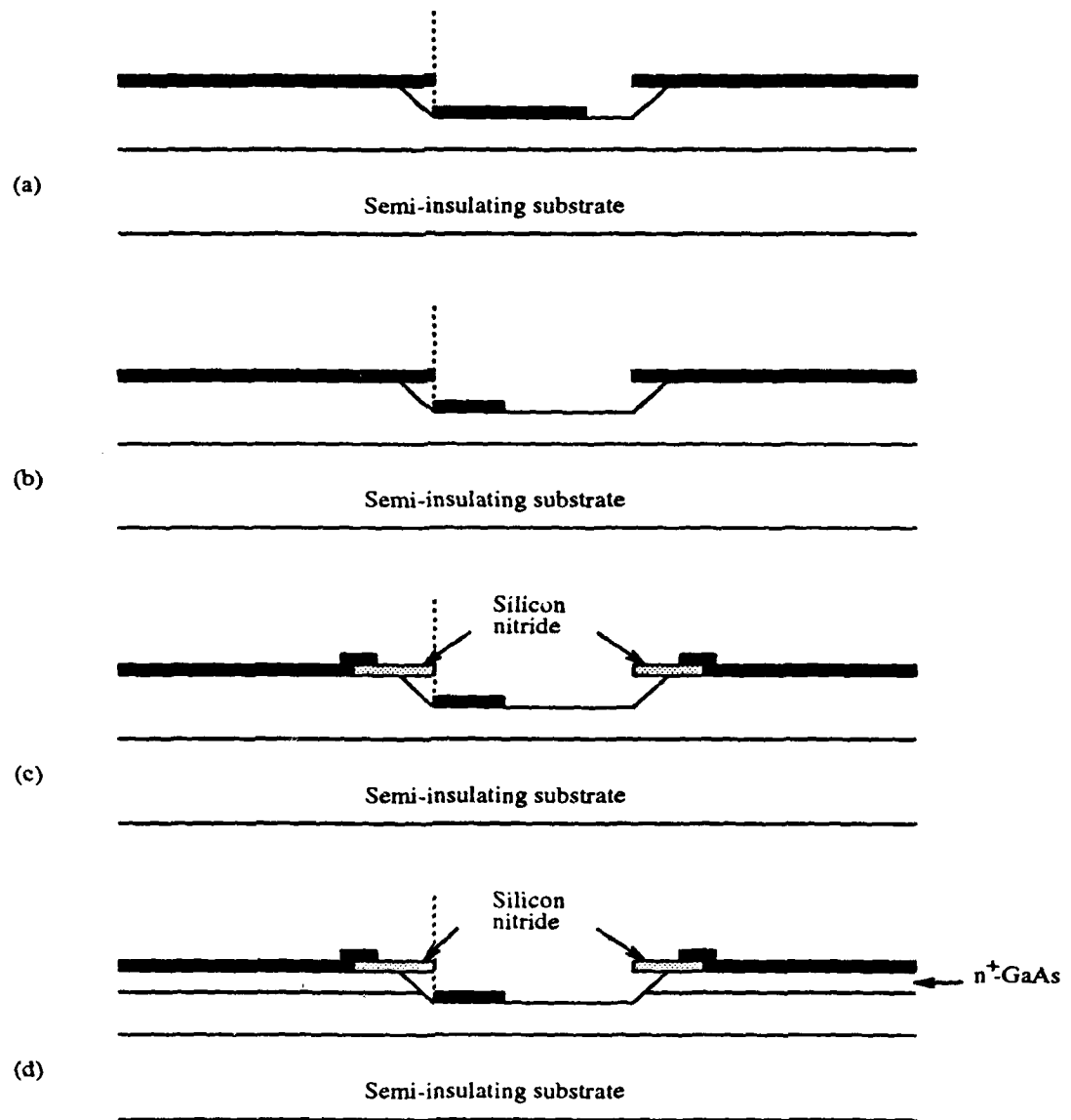
Fig. 3.9 The MESFET structure in which (a) the gate is aligned to the source to decrease the R_{gs}. (b) the separation between the gate and the drain is maximized for increased breakdown voltage. (c) a Si_3N_4 film is inserted to prevent the possible shorting between the gate and the source while still maintaining the self-aligned structure. (d) a n^+ GaAs layer is inserted to again decrease the R_{gs} as well as to facilitate ohmic contacts.
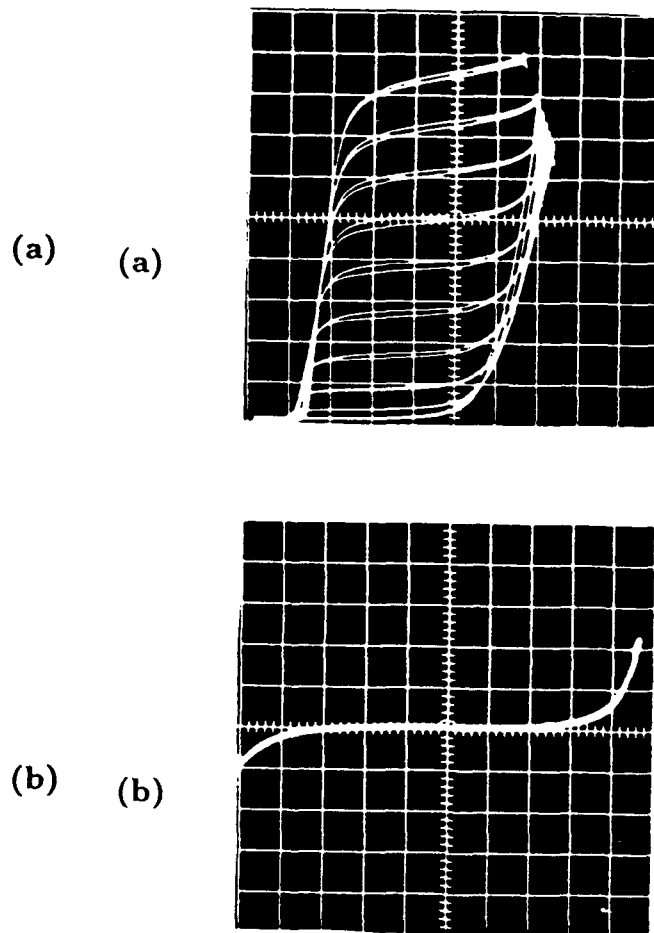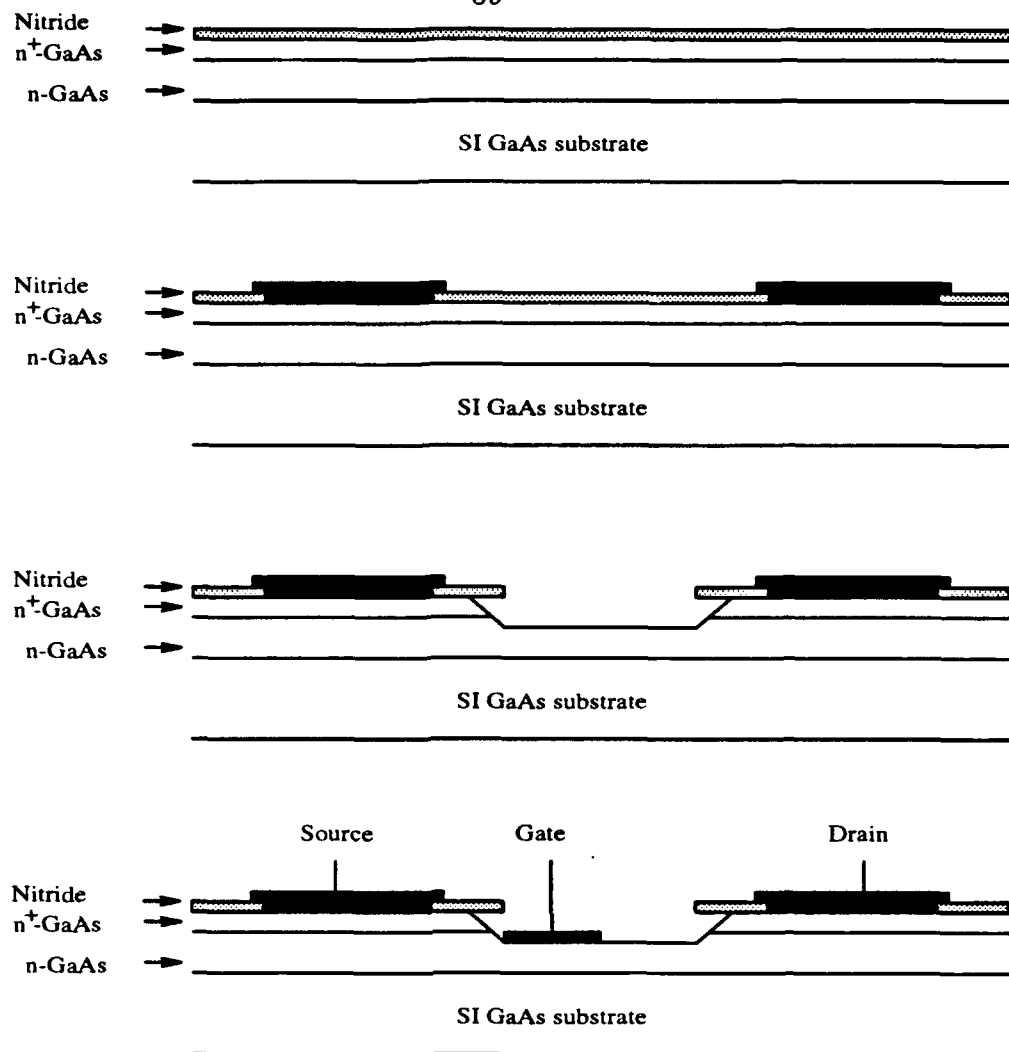
Fig. 3.10 (a) Common-source I-V characteristics of a MESFET showning a break-down voltage of $\approx$ 4 V. The initial turn-on voltage of 1 V is due the LED which is in series with the MESFET. Scales are 500 μA/div vertically and 1 V/div horizontally. (b) The reverse breakdown characteristics of the gate-drain Schottky diode (first quadrant) and the gate-source Schottky diode (third quadrant). Scales are 10 μA/div vertically and 1 V/div horizontally.

of the gate not only achieves a higher breakdown voltage in a MESFET, but also increases the transconductance. This improved structure is seen in Fig. 3.9(b).

Because of the self-aligned nature in defining the gate in a MESFET, it is sometimes inevitable that the gate metalization is shorted to the source metalization due to the close proximity these two metalization are with respect to each other. Therefore, it is necessary to insert a dielectric layer, such as Si_3N_4, which acts as a spacer in preventing the shorting between the gate and the source, and still maintains the self-aligned gate structure with respect to the nitride layer. This is shown in Fig. 3.9(c). The insertion of the Si_3N_4 layer however increases the physical spacing between the gate and the source metalizations, which, in turn, increases the parasitic gate-source resistance, R_{gs}, as mentioned before. Therefore, it is necessary to insert an n^+ GaAs layer beneath the drain and the source metalizations to reduce the actual distance between the source and the gate as well as to decrease the resistance for the drain and the source ohmic contacts. As a result, a MESFET structure shown in Fig. 3.9(d) is obtained. It is a self-aligned and passivated MESFET with a recessed asymmetric gate.

The composition of the material required for this self-aligned and passivated MESFET with a recessed asymmetric gate consists of an n^- GaAs layer beneath an n^+ GaAs layer on a semi-insulating GaAs substrate. The fabrication process of the MESFET is outlined in Fig. 3.11. Firstly, a blank deposition of Si_3N_4 was performed on the wafer followed by etching away the Si_3N_4 at the source and drain ohmic contact regions in a CF_4 plasma. AuGe/Ni/Au were evaporated onto the wafer and lifted off to define the source and the drain. The wafer was then subjected to alloying in a N_2 ambient at 430° C for 4 minutes. The gate recess region was photolithographically defined next and the exposed Si_3N_4 was again etched away in a CF_4 plasma. Once the Si_3N_4 in the gate region was removed, the wafer was immersed in a chemical etchant, consisting of NH_4OH, H_2O_2, and H_2O in a 20 : 7 :

Nitride
n⁺-GaAs
n-GaAs

SI GaAs substrate

Source   Gate   Drain

Nitride
n⁺-GaAs
n-GaAs

SI GaAs substrate

Fig. 3.11 Fabrication steps of a self-aligned and passivated MESFET with a recessed asymmetric gate.
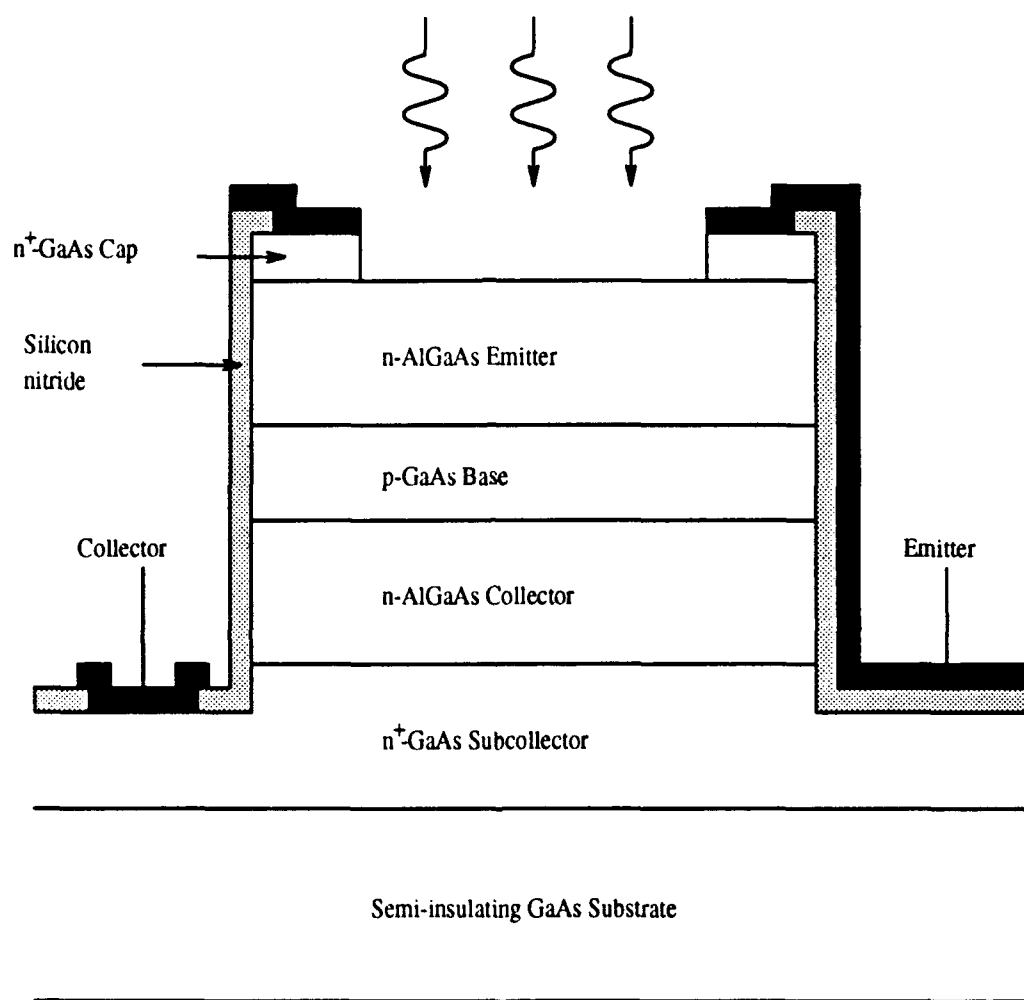
973 ratio, to recess the gate region. While the gate was being recessed, the amount of current flowing between the source and the drain was monitored. Initially, the effect of the etch was probably not apparent from the current measured. However, as the recess became deeper, the source-drain current started to saturate due to the fact the remaining conduction layer began to be pinched off at the drain end. As the etch got even deeper, the saturation current became even smaller. This etching process was continued until the desired saturation current was obtained. Because the gate was to be subsequently evaporated onto this recessed region, there would be an additional depletion region developed underneath the gate due to the gate metalization. This additional depletion region reduced the height of the conduction channel layer, which, in turn, reduced the source-drain saturation current. Therefore, the recessed etching was stopped slightly before the desired source-drain saturation current was reached, so that after the gate metalization the appropriate source-drain saturation current was obtained. Finally, Ti/Pt/Au was evaporated to define the gate in a self-aligned manner as described earlier. The dimensions of the gate was 7×100 μm^2 with a gate to drain spacing of 11 μm.

The I-V characteristics of the MESFET was shown in Fig. 3.10(a) earlier. A transconductance of 30 mS/mm and a source-drain breakdown voltage of 4 V were measured. The initial offset in the V_{DS} was due to the turn-on voltage of the LED, which was in series with the MESFET. These results were consistent with the expectation except for the low breakdown voltage of 4 V. This was probably caused by the surface-induced breakdown instead of the true gate-drain Schottky diode breakdown because any dirt or particles in the vicinity of or underneath the gate would cause the premature breakdown.
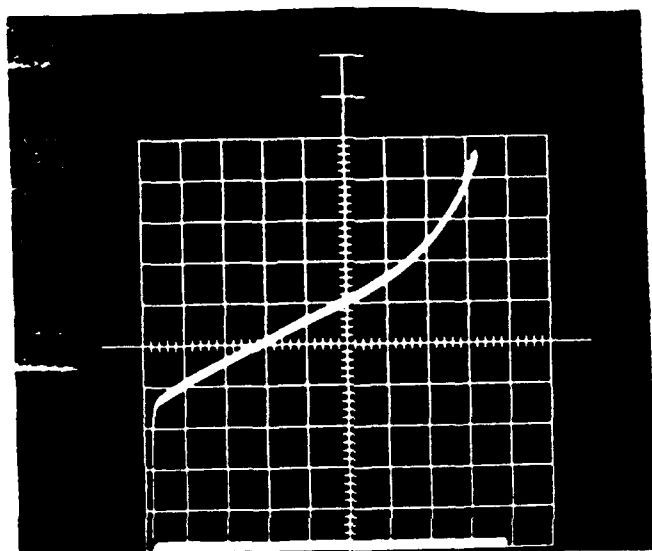
3.3.2 Phototransistors

Phototransistors are bipolar transistors with a floating base. As the light is incident onto the phototransistor window, it traverses through the transparent and high-bandgap emitter region and is absorbed in the small-bandgap base layer. The photocurrent generated acts as the base current and is amplified through the normal amplification process in a heterojunction bipolar transistor to produce the collector current. Since there is an initial efficiency involved in detecting the incoming photons, the overall optical gain of a phototransistor is $\eta_D\beta$, where η_D is the efficiency of generating electrons from the incoming photons and β is the common-emitter transistor current gain. In designing a phototransistor, there is an issue that should be noted. It is the issue of the base layer thickness. A thin base layer should be used to maximize the current gain. However, if the base is too thin, the incoming photons will not be fully absorbed inside the base. Therefore, there is an optimal thickness for the base. Fortunately, the reduction of the current gain of the transistor with a thick base can be compensated by using as more lightly doped base. Consequently, the thickness of the base in the transistor should be chosen first to accommodate the absorption of photons and then to optimize the current gain. There is a disadvantage in using a lightly doped base, however. A lightly doped base causes the base width tc bc modulated by the reverse biased base-collector junction. This modulation results in a reduction in the effective base width, which, in turn, causes the collector current to rise. A rise in the collector current causes the output impedance of the transistor to decrease because the output saturation current is now an increasing function of the emitter-collector voltage. This is the Early effect.

The structure of the phototransistor is shown in Fig. 3.12. It consisted of a lightly p-doped GaAs base layer sandwiched by two higher bandgap n-doped AlGaAs layers, namely the emitter and the collector. n^+ GaAs layers were used for emitter and collector contacts. The side wall of the phototransistor was passivated

n+-GaAs Cap

Silicon
nitride

n-AlGaAs Emitter

p-GaAs Base

Collector

n-AlGaAs Collector

Emitter

n+-GaAs Subcollector

Semi-insulating GaAs Substrate

Fig. 3.12 The structure of a double-heterojunction phototransistor incorporating a
p-doped GaAs layer as the base.

Fig. 3.13 The I-V characteristics of the phototransistor. The intensity of the incoming laser beam is 90 μ W. The scales for the vertical and horizontal axes are 20 μA/div and 2 V/div.

with a Si_3N_4 dielectric film over which the emitter and the collector metalizations ran. The window within which the incoming photons were incident was transparent to the photons by etching away the absorbing n^+ GaAs cap layer. The I-V characteristics of the phototransistor were obtained by monitoring the intensity of a GaAs laser onto the phototransistor and measuring the emitter-collector current simultaneously. This is shown in Fig. 3.13. Because of the low doping concentration used in this phototransistor, there was a severe Early effect, which caused the output impedance of the phototransistor to decrease substantially. From this measurement, an output impedance of 175 KΩ was obtained. Nevertheless, the I-V characteristics of the phototransistor shown in Fig. 3.13 was typical of all phototransistors fabricated. As the intensity of the laser beam increased, the current level increased as well. For this particular measurement, the input laser beam intensity was 90 μW and measured current was approximately 90 μA at a collector-emitter voltage of 4 V. This corresponded to an external efficiency of 1 A/W. Assuming an absorption efficiency of 0.3 A/W, we obtained a current gain, β, of only 3. This was a result of having a very thick base layer, which was 1.5 μm. The breakdown voltage of the phototransistor was 20 V, indicating the effectiveness of using a high bandgap and lightly doped AlGaAs collector layer.

3.3.3 Optical Field-Effect Transistors

Another candidate for the detector is the optical FET. It is essentially identical structurally with the conventional MESFET except the optical FET does not have a gate. Instead, it uses an optical input, which is incident on the gate area, to control the channel current underneath. There are two possible mechanisms in explaining the operation of an optical FET. One is based on MESFET-like mechanism [27]. As the electron-hole pairs are excited by the incoming laser beam,

a portion of the electrons flow to the surface depletion region, in which the ionized donors are positively charged. This changes the surface potential with respect to the channel potential. As a result, the channel current is modulated by the transconductance of the underlying MESFET, which leads to an increase in the output source-drain current. However, Gummel et al. [28,29] have argued for a different mechanism, which is based on photoconductivity. As the electron-hole pairs are generated by the laser beam, external carriers are injected into the bulk of the photo-excitation region through the source and drain ohmic contacts in order to satisfy the steady-state recombination and generation requirement. In doing so, at rate at which these external carriers are injected into the photo-excitation region depends upon the carrier recombination lifetime and the device transit time. The longer the carrier recombination lifetime is, the more carriers are injected because the recombination rate is the ratio of the carriers injected over the carrier lifetime. Thus, longer carrier lifetime leads to a large carrier injection per unit time, which is current. However, on the other side, these injected carriers are swiftly removed by the electric field inside the device so that the there will be no carrier build-up over time. The faster the carriers are removed from the device, the larger the current is. Thus, the measured current due to the photo-excitation is expected to be inversely proportional to the carrier transit time across the device. In Sec. 3.5, a detailed analysis of the gain mechanism of a photoconductor is presented. In that analysis, it is found that the optical gain of a photoconductor is given by the ratio of carrier recombination lifetime over the carrier transit time. Thus, in maximizing the gain from the optical FET, the gap between the source and the drain should be minimized and yet should still allow sufficient input light to be detected. Of these two mechanisms, the photoconductivity is the more likely explanation for the operation of the optical FET. This is based on the observation that some of the gain measured [28] is too high to be attributable to MESFET-like amplification.
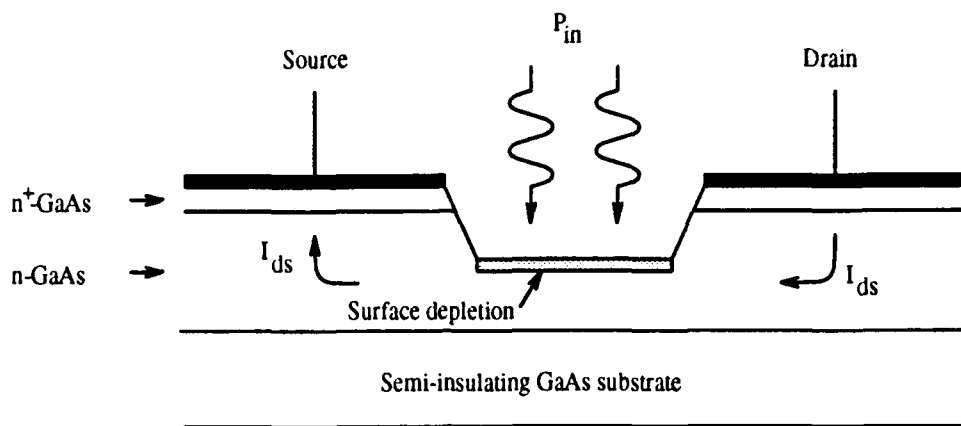
(a)
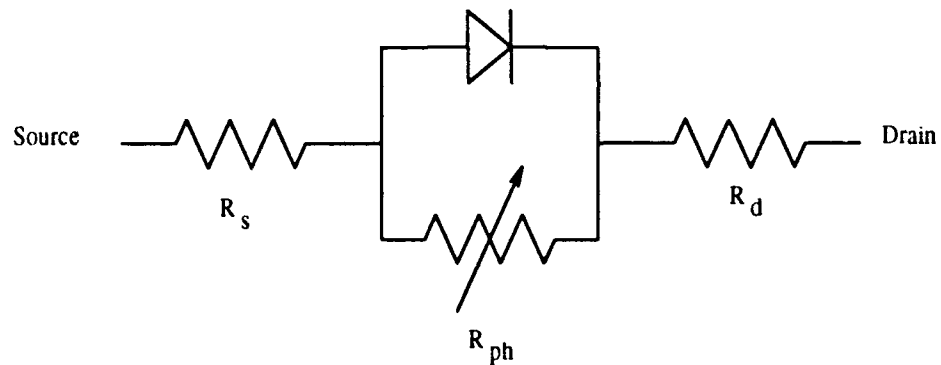
(b)

Fig. 3.14 (a) The cross sectional view of an optical FET. The measured source-drain current is a combination of the p-n photodiode current and the photoconductor current. (b) The circuit model of the optical FET.

Figure 3.14(a) shows the cross sectional structure of an optical FET. Since it is very similar to the conventional MESFET, the fabricational procedure of an optical FET is identical to that of a MESFET, except the gate metalization is not defined. Instead, the amount of the recessed depth is used to control the sensitivity and dark current desired. Generally speaking, the deeper the recess is, the smaller the dark current becomes and the smaller the optical gain is. To explain this dependency, a circuit model shown in Fig. 3.14(b) is proposed to model the optical FET assuming it is dominated by the photoconductivity mechanism. The current channel region is modeled as a reverse biased photodiode in parallel with a photoconductor. The existence of the photodiode is attributed to the fact that there is a surface depletion layer on the exposed surface of the recessed region. Carriers generated in this region are collected by the build-in electric field in the depletion region. Thus, the principle is the same as that of a conventional reverse biased p-n photodiode. Underneath this surface depletion layer, there lies a undepleted ohmic conduction channel made out of n-GaAs. Carriers absorbed in this region contribute to the photoconductivity action as described earlier. To complete the modeling, the source and the drain contact and bulk resistances are added in series. Since the photoconductor is a high-gain detector and the photodiode does not have any gain, the overall efficiency of the optical FET would be bounded by the efficiency of these two devices. If the photoconducting channel region is thick, a high-gain optical FET can be expected. As this region gets thinner by the recessed etch, the photoconductivity effect starts to decrease due to the smaller absorption region. As a result, the optical gain decreases gradually. This process continues until this photoconducting channel totally disappears, leaving only the surface depletion region. At this point, the optical FET does not exhibit any optical gain because the remaining photodiode does not have any gain. These characteristics are observed in actual experimental data, which are shown in Fig. 3.15 and 3.16.
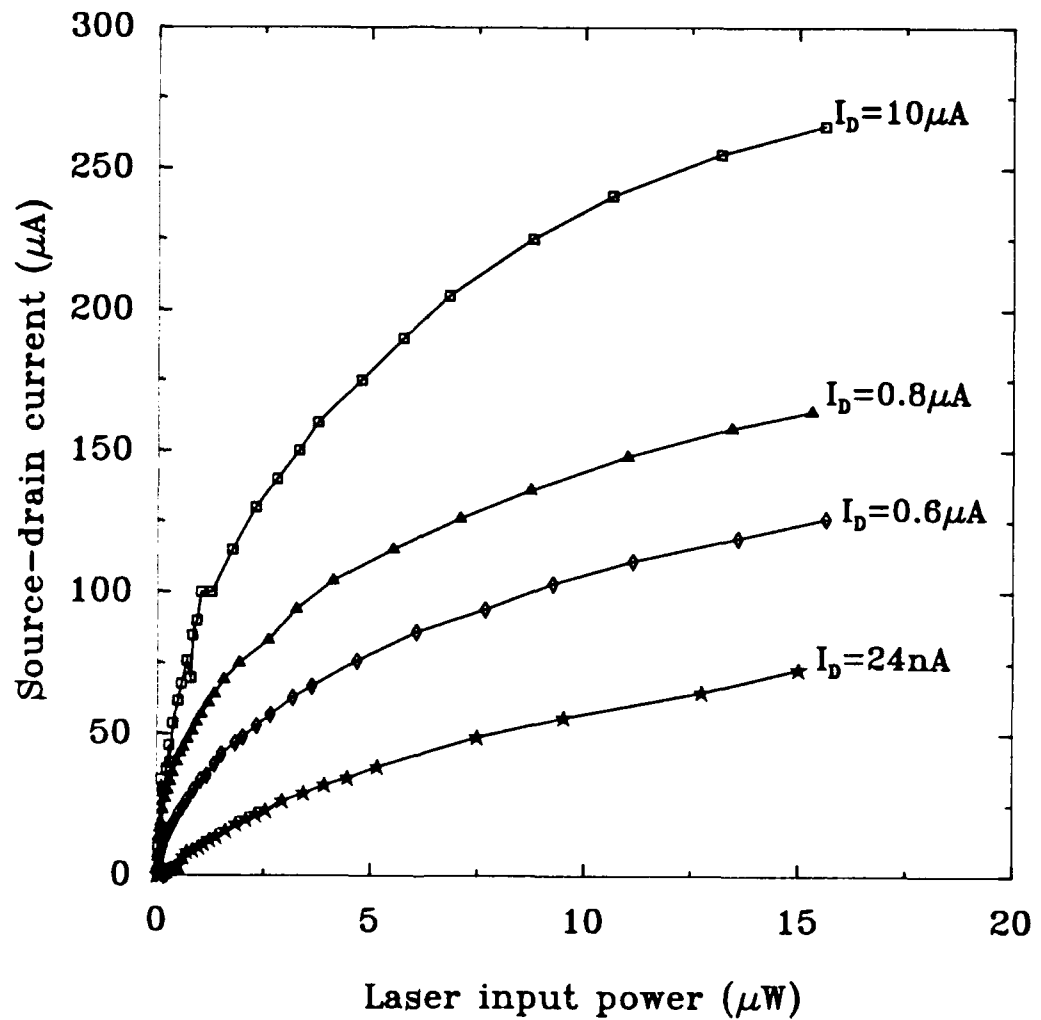
Fig. 3.15 Input-output characteristics of an optical FET. The output is the measured source-drain current and the input is the intensity of the laser beam. The four curves correspond to four different dark currents, which are due to the different recessed depth.
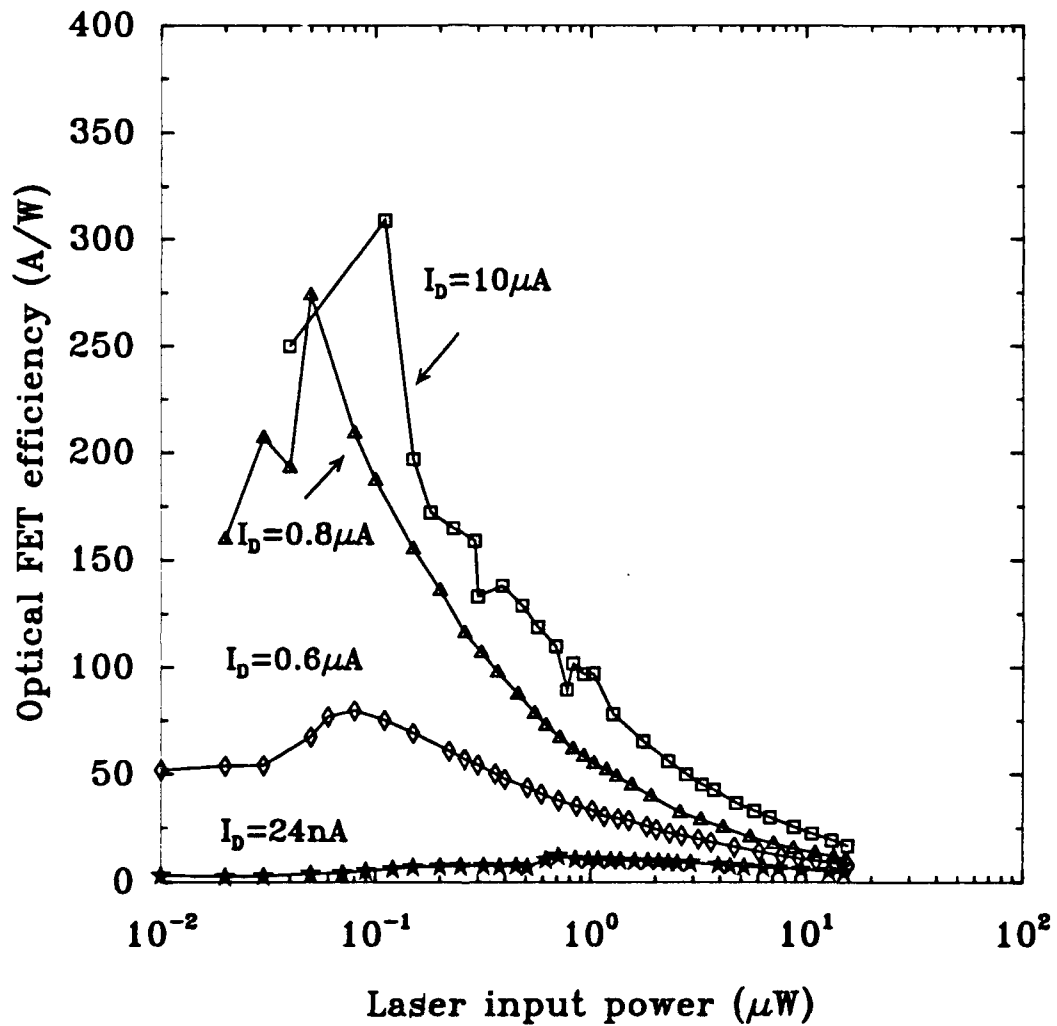
Fig. 3.16 Efficiency of the optical FET plotted in term of A/W as a function of the input power. The vertical axis is re-plotted from Fig. 3.15 by taking the slope of the curve. The horizontal axis remains the same.

Figure 3.15 shows the measured source-drain current as a function of the input laser beam power for four different dark currents, which correspond to four different recessed depths. The higher the dark current, the shallower the recess is. It is clear from this figure that the current increases monotonically as a function of the input power. However, there is a slight saturation in the current as the input power increases. Moreover, as the recessed depth gets larger, the measured current decreases. This is consistent with the argument presented earlier in the modeling of the optical FET. Figure 3.16 shows the efficiency of the optical FET in A/W for the same four sets of measurement with the same dark current. The efficiency of the optical FET increases initially as the input power increases. However, it decreases as the input power continues to increase. The mechanism of this unexpected behavior is not clear. But, this trend is consistent within each curve. However, the drop in efficiency of the optical FET as the recessed depth increases is in agreement with the model presented earlier.

3.4 Experimental Results of MESFET-Based Neurons

Having discussed the design consideration and analysis of the MESFET-based neuron as well as the individual devices in the neuron, we next describe the process of fabricating the optoelectronic neuron and the testing results. Figure 3.17 shows the device cross sectional view of the monolithically integrated optoelectronic ne urons consisting of the device elements shown in Fig. 3.1. Basically, on the semi-insulating GaAs substrate, an undoped GaAs buffer layer was first grown. Upon which, an n^+-GaAs acting as the source and the drain ohmic contact layer on top of an n^--GaAs current conduction channel layer were grown. These two layers form the structure of the MESFET. On top of the n^+-GaAs layer, a conventional double heterojunction bipolar transistor structure was grown. It consisted of an n^+-GaAs

layer as the subcollector layer, an n-$Al_{0.35}Ga_{0.65}As$ layer as the collector layer, a p-GaAs layer as the base layer, an n-$Al_{0.35}Ga_{0.65}As$ layer as the emitter layer and an n^+-GaAs layer as the contact layer. The doping concentration and the thickness of each layer are listed in Fig. 3.18. The formation of the LED was completed by diffusing Zn twice over different areas to create the confinement for the current.

The fabrication of the MESFET-based optoelectronic neuron began by applying the standard degreasing and cleaning procedure to the surface of the GaAs epitaxial layers. A non-selective etch, consisting of a mixture of H_3PO_4 , H_2O_2 and CH_3COOH in the ratio of 1 : 1 : 3 was used to etch down to the n^+-GaAs layer to define the LED and the phototransistor. After this, the same etch was used to etch down the semi-insulating substrate to define the MESFET. A blank deposition of Si_3N_4 was then applied to the surface of the device by using a thermal chemical vapor deposition system heated to 610°. The gases used were silane diluted to 1% by nitrogen, ammonia and nitrogen. A thickness of approximately 1200 Å to 1500 Å of Si_3N_4, which exhibited a color of blue to light blue. was deposited. The next step was Zn-diffusion to convert the n-AlGaAs emitter layer to p-AlGaAs for the upper cladding layer for the LED as well as to provide the current confinement. This was achieved by selectively removing the Si_3N_4 over the the LED window area in a CF_4 plasma and performing a sealed ampoule Zn-diffusion process. The ampoule, in which the neuron device and the diffusion source, $ZnAs_2$ were placed, was pumped to a vacuum of 8×10^{-8} torr before it was sealed with a torch. The ampoule was then inserted into a furnace of 640° to promote the diffusion of Zn into the exposed area of the LED for approximately 9 minutes. Afterwards, the ampoule was quickly quenched. Usually there was As condensation on the inner wall of the ampoule after quenching to indicate the proper diffusion of Zn into the LED. A second diffusion process was carried out by using exactly the same procedure except the diffusion time was approximately 5 minutes and the area of diffusion
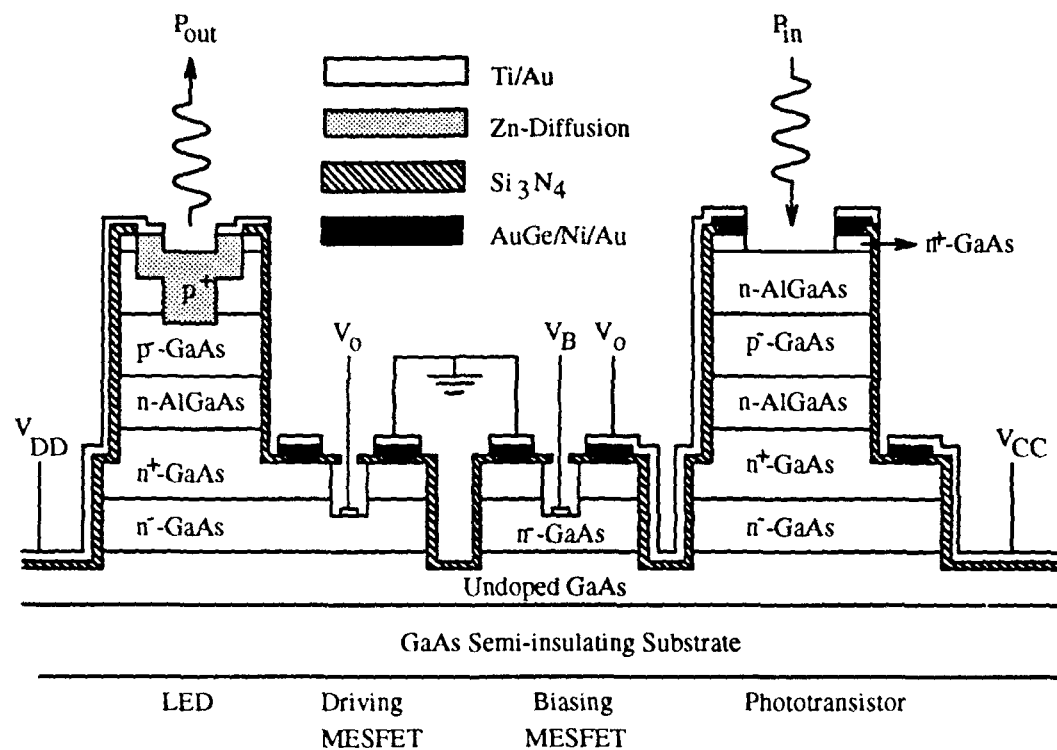
Fig. 3.17 The cross sectional view of the MESFET-based optoelectronic neuron monolithically integrating 2 MESFET's, a LED and a phototransistor.

GaAs	$n = 5 \times 10^{18}$ cm^{-3}	0.15 μm
Al$_{0.35}$Ga$_{0.65}$As	$n = 2 \times 10^{17}$ cm^{-3}	0.3 μm
GaAs	$p = 1 \times 10^{16}$ cm^{-3}	1.5 μm
Al$_{0.35}$Ga$_{0.65}$As	$n = 5 \times 10^{16}$ cm^{-3}	0.6 μm
GaAs	$n = 5 \times 10^{18}$ cm^{-3}	1 μm
GaAs	$n = 2 \times 10^{17}$ cm^{-3}	0.25 μm
GaAs	undoped	0.4 μm

Semi-insulating GaAs Substrate

Fig. 3.18 The epitaxial material composition of the MESFET-based optoelectronic neuron.

n$^+$-GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n$^+$-GaAs
n$^-$-GaAs
undoped GaAs

Semi-insulating GaAs substrate

LED/Phototransistor definition

n$^+$-GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n$^+$-GaAs
n$^-$-GaAs
undoped GaAs

Semi-insulating GaAs substrate

MESFET definition

n$^+$-GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n$^+$-GaAs
n$^-$-GaAs
undoped GaAs

Semi-insulating GaAs substrate
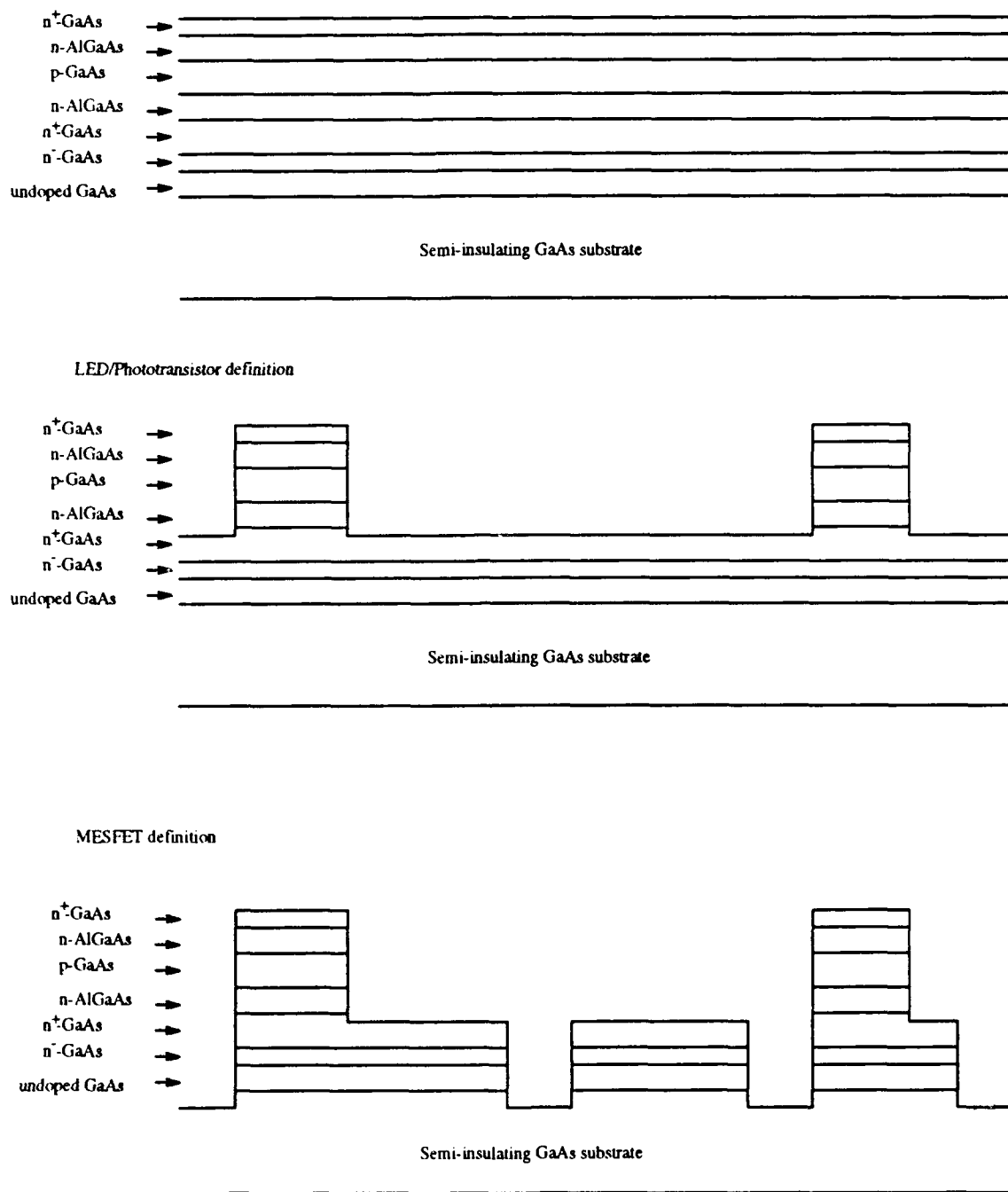
Fig. 3.19(a) The fabrication steps of the MESFET-based optoelectronic neuron.

Silicon nitride deposition

Silicon nitride

n -GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n⁺-GaAs
n⁻-GaAs
undoped GaAs

Semi-insulating GaAs substrate

Silicon nitride

First Zn diffusion

Zn diffusion

n⁺-GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n⁺-GaAs
n⁻-GaAs
undoped GaAs

Semi-insulating GaAs substrate

Silicon nitride

Second Zn diffusion

Zn diffusion

n⁺-GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n⁺-GaAs
n⁻-GaAs
undoped GaAs

Semi-insulating GaAs substrate

Fig. 3.19(b) The fabrication steps of the MESFET-based optoelectronic neuron.

76

Nitride opening

Silicon nitride
Zn diffusion

n+-GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n+-GaAs
n--GaAs
undoped GaAs

Semi-insulating GaAs substrate

N-type metalization/Alloy

Silicon nitride
Zn diffusion
AuGe/Ni/Au

n+-GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n+-GaAs
n--GaAs
undoped GaAs

Semi-insulating GaAs substrate

Gate recess etch

Silicon nitride
Zn diffusion
AuGe/Ni/Au

n+-GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n+-GaAs
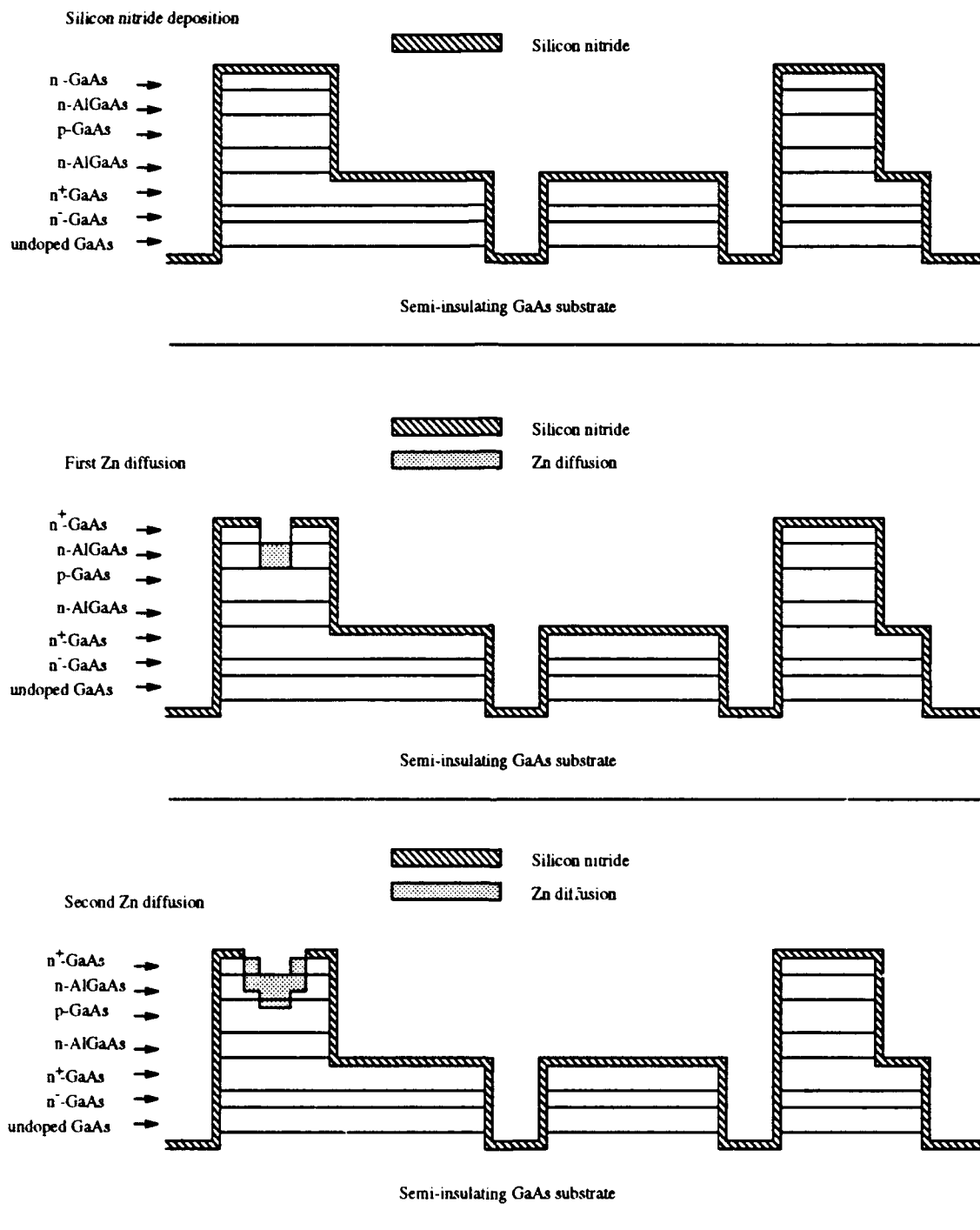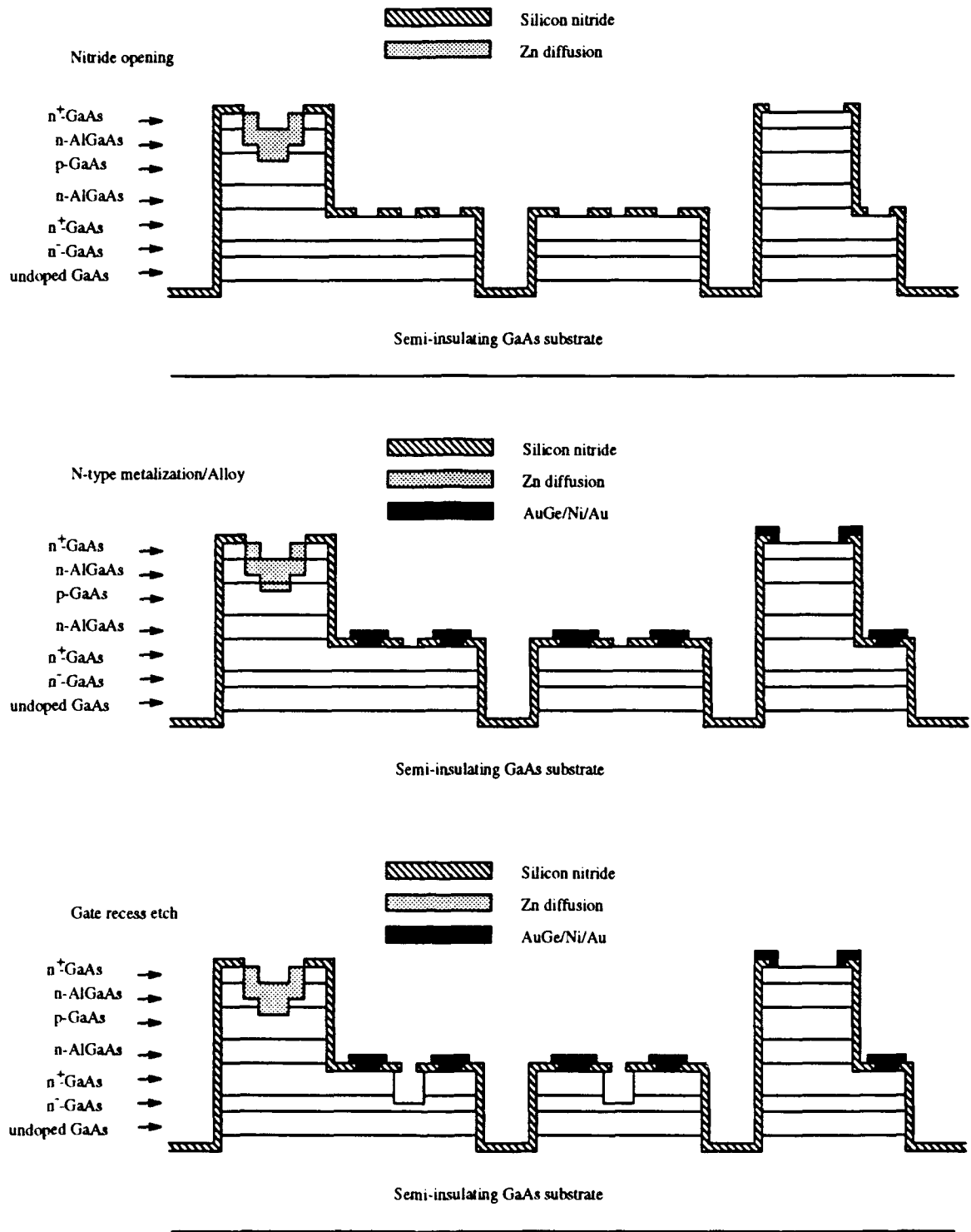n--GaAs
undoped GaAs

Semi-insulating GaAs substrate

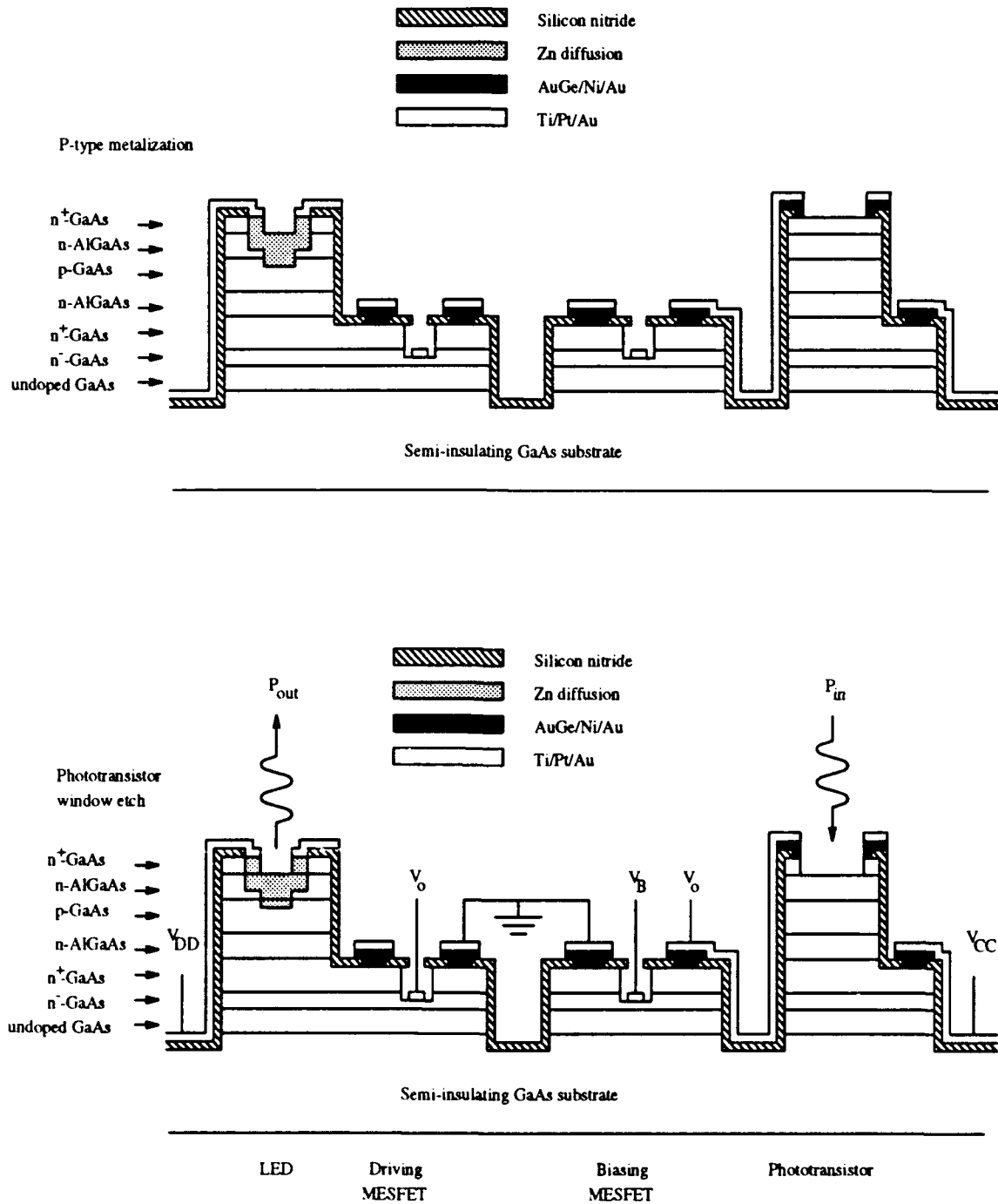**Fig. 3.19(c)** The fabrication steps of the MESFET-based optoelectronic neuron.

Fig. 3.19(d) The fabrication steps of the MESFET-based optoelectronic neuron.

was slightly larger. After the Zn-diffusion step, selective area of Si_3N_4 was again removed in CF_4 plasma to facilitate the subsequent ohmic contacts for the source and the drain as well as the gate recessed area of the MESFET, and the contacts for the emitter and the collector of the phototransistor. The transistor contact terminals, including the source and the drain, and the emitter and the collector of the phototransistor, were metalized by evaporating AuGe/Ni/Au of $200\mathring{A}$, 100 $\mathring{A}$, and 1500 $\mathring{A}$, respectively, by using the lift-off technique and subsequently alloyed at 430 °C in an N_2 ambient for 4 minutes to drive in the Ge in forming the ohmic contacts. The gate recess etching process was then performed by monitoring the source-drain current in the MESFET. This etching used the Si_3N_4 as the mask in order to obtain a self-aligned recess. The etching was stopped when the desired source-drain saturation was obtained. The etchants used in recessing the gate was NH_4OH, H_2O_2 and H_2O in a ratio of 20, 7 and 973 respectively. The etch rate was approximately 30 $\mathring{A}$/second. Next, the gate was defined by evaporating first 150 $\mathring{A}$ of Ti and then 1500 $\mathring{A}$ of Au by an electron beam evaporator. The excess metals were lifted off in acetone. The area of the gate was 6×100 μm^2 and was self-aligned asymmetrically to the edge of the source inside the recessed region. The last step was to remove the light-absorbing n^+-GaAs cap layer in the phototransistor by wet etching. The entire area, including the contact pads, measured approximately 400 $\times$ 400 μm^2. However, the active device area was only about 150×250 μm^2.

The neuron was tested by illuminating the phototransistor window area by a GaAs laser diode. This was achieved by splitting the output beam of the laser diode by a beam splitter into 2 equal-intensity beams. One of the beams was focused onto the phototransistor and the other beam was focused into the detector in order to monitor the power of the beam incident on the phototransistor. The input switching circuit was first tested. The voltage between the phototransistor and the loading MESFET, V_{DS1}, was monitored as the intensity of the input laser

beam was varied. A power supply of 2.5 V was connected to the collector of the phototransistor, while the source of the MESFET was electrically grounded. By varying the gate voltage of the MESFET, V_{DS1} was measured at a fixed laser output power incident on the phototransistor. Then the laser output power was changed and the same measurement was performed again. The results are shown in Fig. 3.20, in which V_{DS1} is plotted against V_B for five different laser powers incident on the phototransistor. For a laser power of 10.8 μW incident on the phototransistor, the voltage, V_{DS1}, was pulled up to the power supply voltage less the phototransistor saturation voltage. As the gate voltage of the MESFET increased, V_{DS1} stayed relatively unchanged until the current drawn by the MESFET had exceeded the photocurrent provided by the phototransistor. At which point, V_{DS1} dropped and was pulled down to ground. As the laser power became smaller, the value of the gate voltage at which V_{DS1} dropped from 2.5 V to ground decreased. This was consist with the analysis shown earlier because as the photocurrent became smaller, the current needed by the loading MESFET to pull down V_{DS1} also became smaller. However, due to the leakage current from the drain to the gate of the MESFET, V_{DS1} could not be pulled up completely. In fact, as the gate bias decreased, this leakage current increased because the gate-drain had became more reverse biased. Since this leakage current was no different in nature compared to the source-drain current drawn by the MESFET from the standpoint of V_{DS1}, V_{DS1} was pulled down as a result. This imperfection was clearly evidenced for laser input power of 3.2 μW and less. Thus, it was extremely important that the leakage current be minimized in the MESFET, especially the one across the gate-drain terminals.
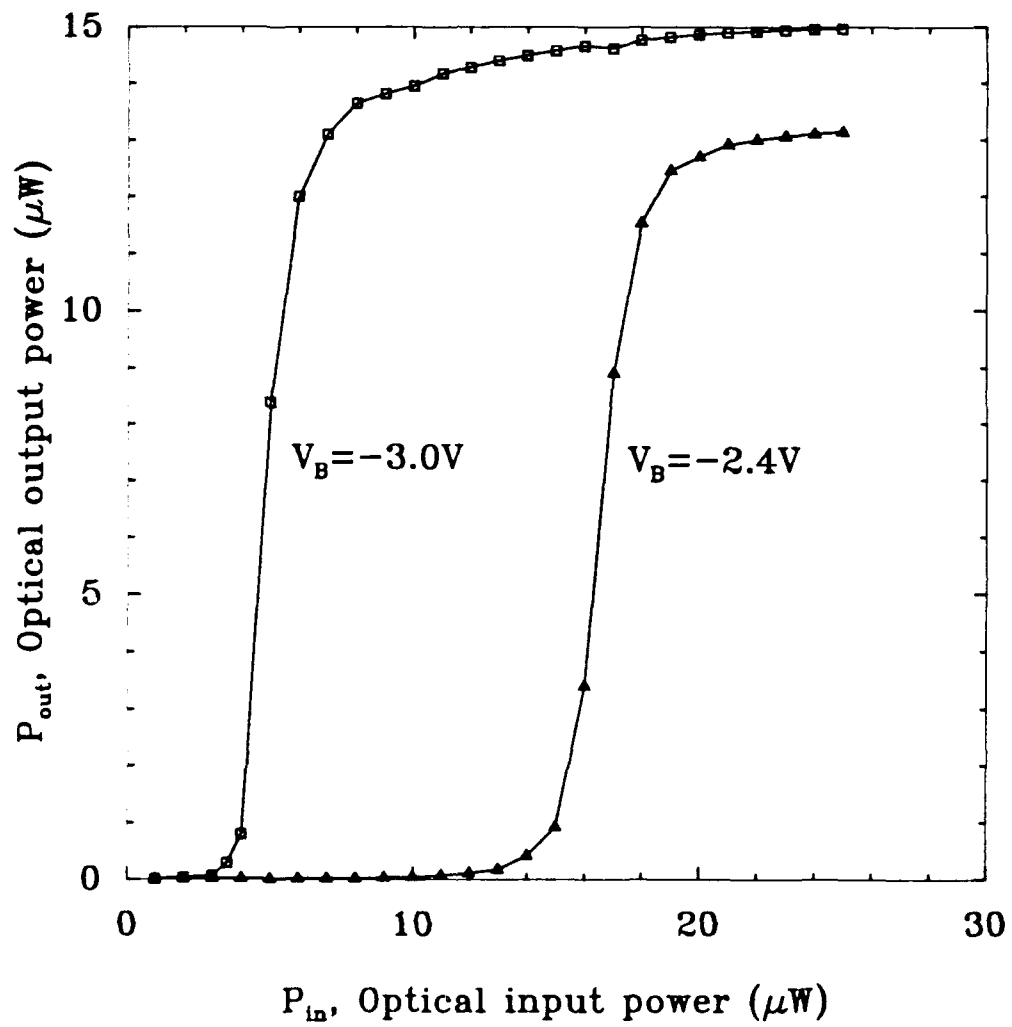
The overall input-output characteristics of the optoelectronic neuron was obtained by monitoring the current through the LED as a function of the laser input power incident on the phototransistor at a fixed gate voltage on the loading MESFET. Figure 3.21 shows two of these plots. One of them was taken at a gate voltage
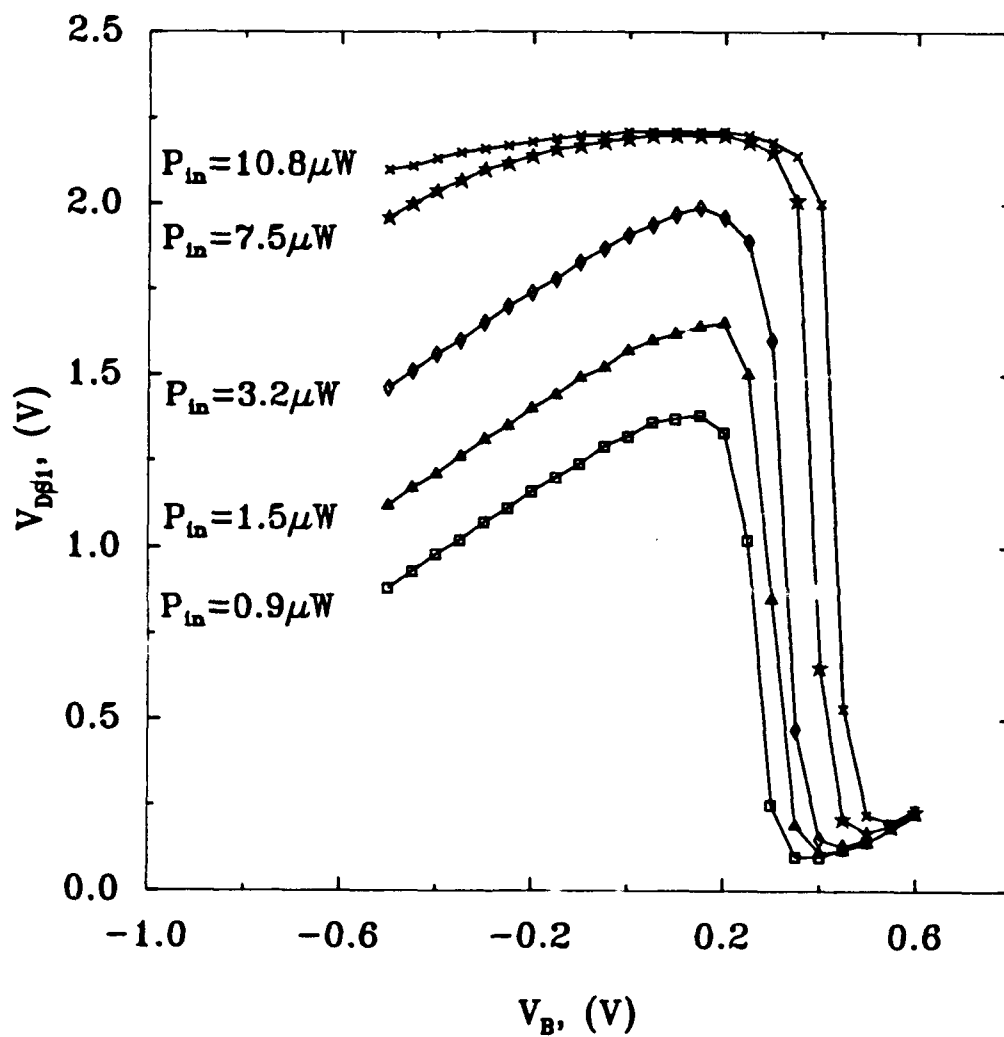
Fig. 3.20 V_{DS1} as a function of the gate voltage on the loading MESFET for various input laser power incident on the phototransistor.

of -3.0 V and the other one was taken at a gate voltage of -2.4 V. The measured LED current was converted into optical output power by assuming an external quantum efficiency of the LED to be 0.01 W/A. The reason the output power of the LED could not be directly measured was because the beam of the LED diverged too fast and it was difficult to collect all of it into the detector. If, however, we brought the detector very close to the neuron, the input laser beam could not easily illuminate the phototransistor. Thus, the current through the LED was monitored. The external efficiency of 0.01 W/A was typical for the LED of double Zn-diffusions. For the curve with a gate voltage of -3.0 V, the output remained zero until the input power reached approximately 3 μW. Beyond this point, the output power increased rapidly to 12 - 15 μW over an additional input of 2 μW. This represented a differential optical gain of 6. The threshold of the neuron was controlled by applying a different voltage to the gate of the loading MESFET, as clearly seen in the plot. Because of the leakage in both the loading MESFET and the output driving MESFET, a minimum 3 μW was necessary to turn on the neuron. By reducing the leakage currents through the gate in both MESFET's, this number is expected to drop substantially. During the on-state of the neuron, the LED current was measured to be 1.2 mA. For a 2-Volt power supply on the output driving circuit, the electrical power dissipation was 2.4mW. When the input laser beam was pulsed to a level just enough to turn on the neuron, the output LED current showed a rise time of 5 μsec. This is shown in Fig. 3.22. This meant that the neuron could be turned on with an optical switching power of only (2 μW) $\times$ (5 μsec) = 10 pJ [30]. Not only did this MESFET-based optoelectronic neuron exhibit comparable switching energy as compared to the SEED devices [31], it also dissipated only 2.4 mW of electrical power. This was a factor of 40 less when compared to the DHBT-based neuron.

The differential optical gain of 6 was limited by the finite output impedance of

**Fig. 3.21** Input-output characteristics of the MESFET-based optoelectronic neuron. V_B is the bias voltage on the gate of the input switching circuit.
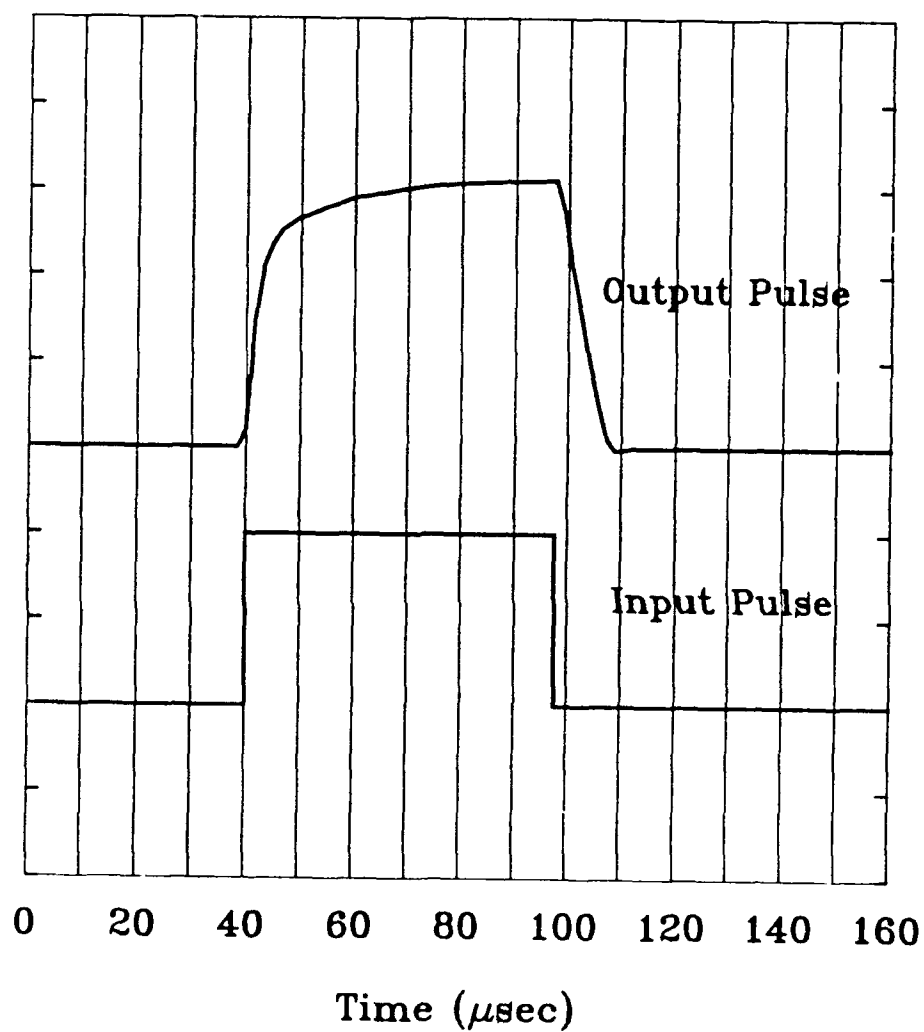
Fig. 3.22 Time measurement of the MESFET-based optoelectronic neuron in response to a step input in the laser power incident on the phototransistor. The rise time was measured to be 5 μsec.

the phototransistor and the loading MESFET as well as the leakage currents in the MESFET's. The output impedance of the phototransistor could be increased by doping the base more heavily. However, this reduced the current gain of the photo-transistor. As a result, the base thickness had to be reduced to compensate for the increased doping concentration in order to maintain the same current gain. Unfortunately, reducing the thickness of the base layer adversely affected the absorption efficiency of the phototransistor. Therefore, an optimized design had to be used. The output impedance of the loading MESFET could be increased by using a more insulating substrate as well as reducing the leakage current through the gate. It was interesting to note that reducing the leakage current has a lot of benefits in terms of improving the optical gain and the sensitivity of the neuron. Thus, the same MESFET-based optoelectronic neuron was fabricated again by carefully cleaning the surface before the gate metalization was defined. Furthermore, a different gate metalization composition was employed. This consisted of the same Ti/Au metals except an 100-$\overset{\circ}{A}$ layer of Pt was inserted between the Ti and the Au. The doping concentration of the MESFET conduction layer was also reduced to 5×10^{16} cm^{-3} for less leakage current across the gate. Figures 3.23 - 3.27 show the results of the neuron, which incorporated the above-mentioned simple changes [32].

First of all, a comparison of the gate leakage current in the old and the new neurons was made. Figure 3.23 shows the measured gate-drain reverse leakage current as a function of the reverse biased voltage for the two Schottky diodes. At a typical operating voltage of 1 volt across the gate and the drain, the leakage current for the new MESFET was at least an order of magnitude lower. This reduction in the leakage current directly translated into an increase in the optical gain of the neuron because the phototransistor needed a smaller photocurrent to pull up the gate voltage of the output driving MESFET. Figure 3.24 shows the input-output relationship of the improved MESFET-based optoelectronic neuron. The
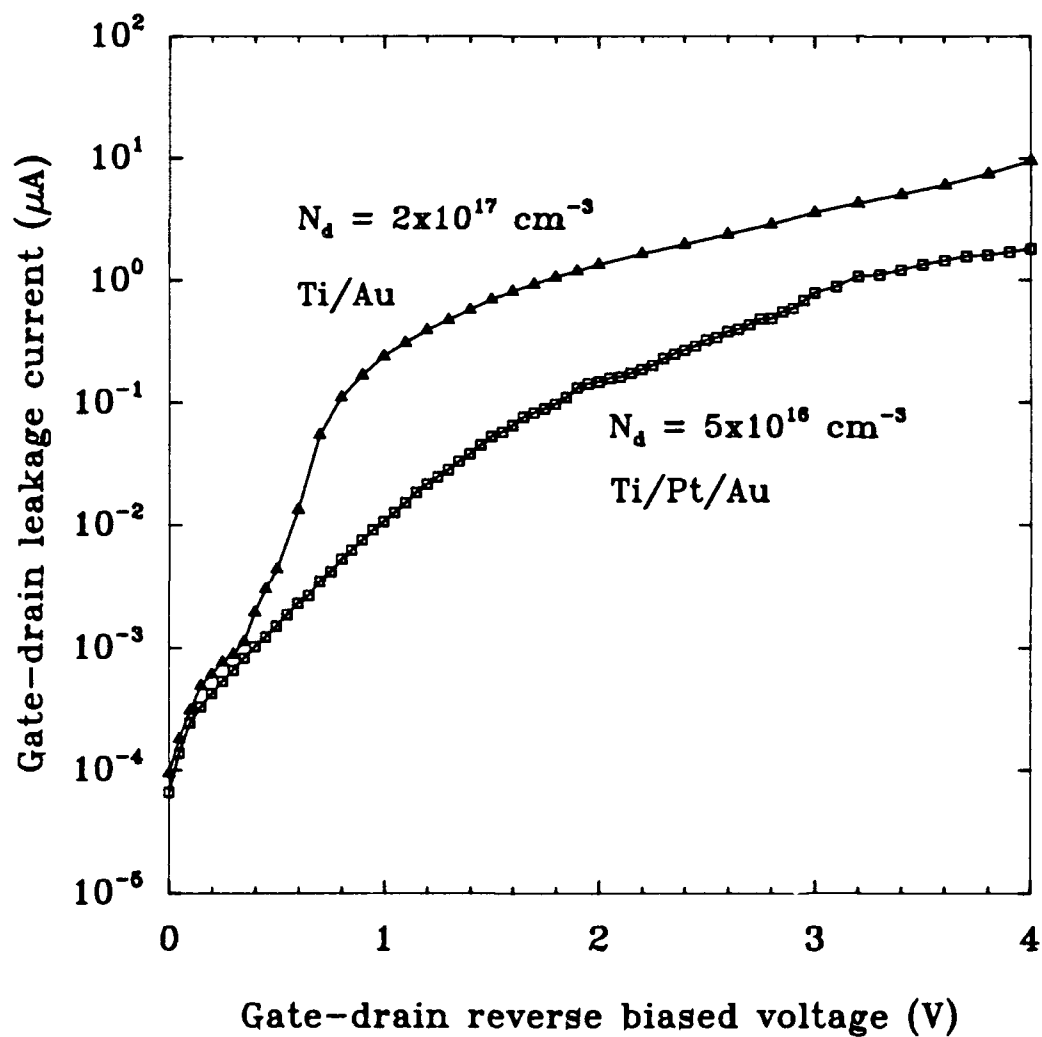
Fig. 3.23 Comparison of the gate-drain leakage currents as a function of the gate-drain reverse biased voltage in the loading MESFET for the old and the new neurons.

testing conditions were the same as before except the gate of the loading MESFET was floating. This was intended to reduce the gate-drain leakage current further. It is evident from the plot that, by reducing the gate-drain leakage current, the minimum input power needed to turn on the neuron was reduced. In this case, an 1 μW input power was measured. Moreover, the differential optical gain also increased dramatically to 40 as an additional input power of 0.2 μW beyond the threshold caused a change of 8 μW at the output. This improvement was remarkable considering the only improvement made was to reduce the gate leakage current of the loading MESFET. Not only was the differential optical gain improved, but also the absolute optical gain had increased to 8. The current drawn by the LED during the on-state of the neuron was measured to be 0.8 mA. Therefore, the electrical power dissipation per neuron was 1.6 mW by using a 2-volt power supply. The speed of the neuron was also measured by applying an electrical pulse to the laser diode that illuminated the phototransistor. Figure 3.25 shows the measured response of the neuron. A rise time of 65 μsec was obtained in this neuron. This implied a total optical switching energy of (65 μsec $\times$ 0.2 μW) = 13 pJ. This optical switching energy was comparable to that of the previous neuron, which was 10 pJ. This is expected because the total charges needed to charge up the gate of the output driving MESFET's, which had the same gate width and length, in both neurons were the same. Since the voltage swings at the same gate from the off-state to the on-state of the neuron were also the same, the switching energy, which was equal to QV, remained unchanged. Thus, overall, the neuron became more sensitive and provided more gain. However, this was achieved at the expense of a lower switching speed.

As the leakage current problem was improved, the limiting factor in the performance of the MESFET-based optoelectronic neurons shifted to the efficiency of the detector, which, in this case, was the double-heterojunction bipolar phototransistor.
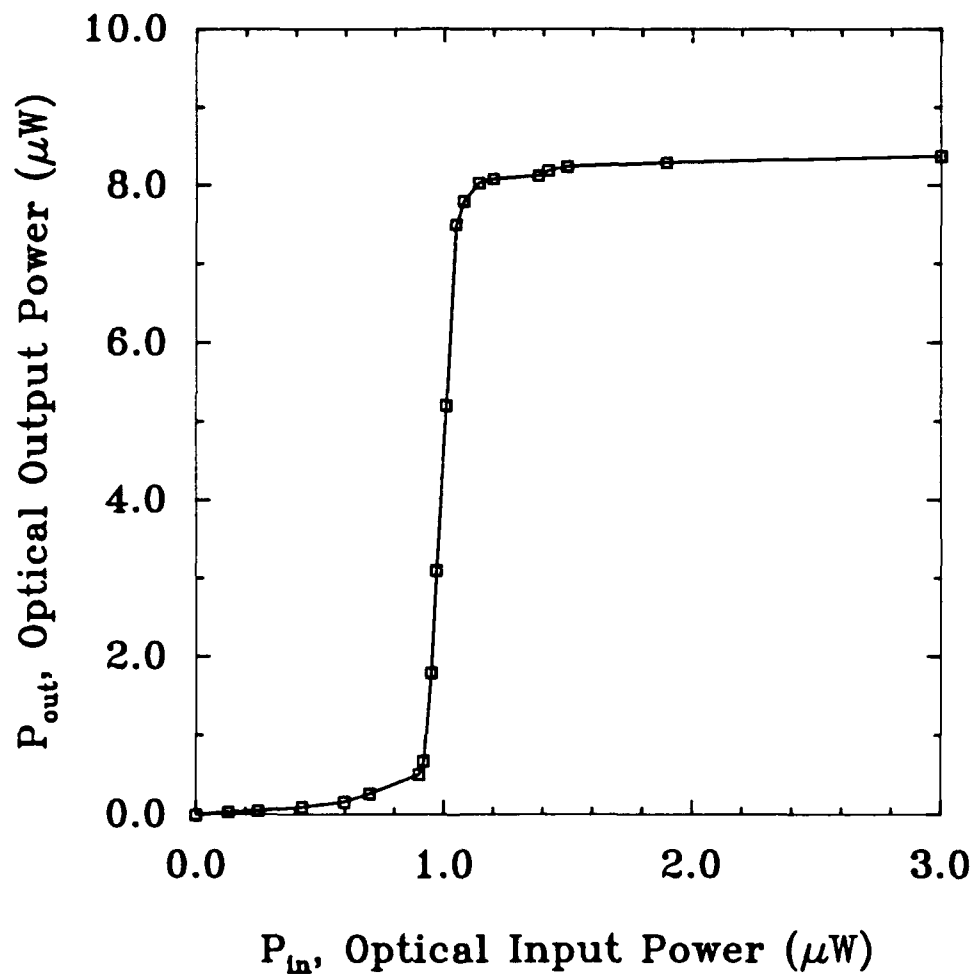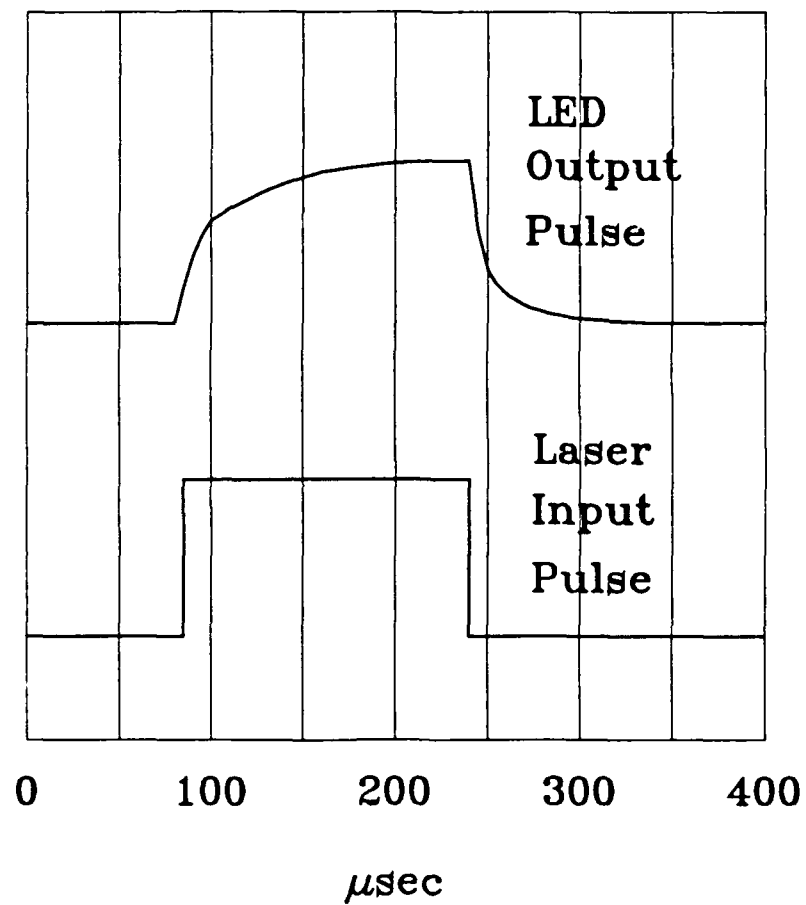
Fig. 3.24 Input-output characteristics of the improved MESFET-based optoelec-tronic neuron, showing a differential optical gain of 40 and an absolute optical gain of 8.

Rise Time = 65 μsec

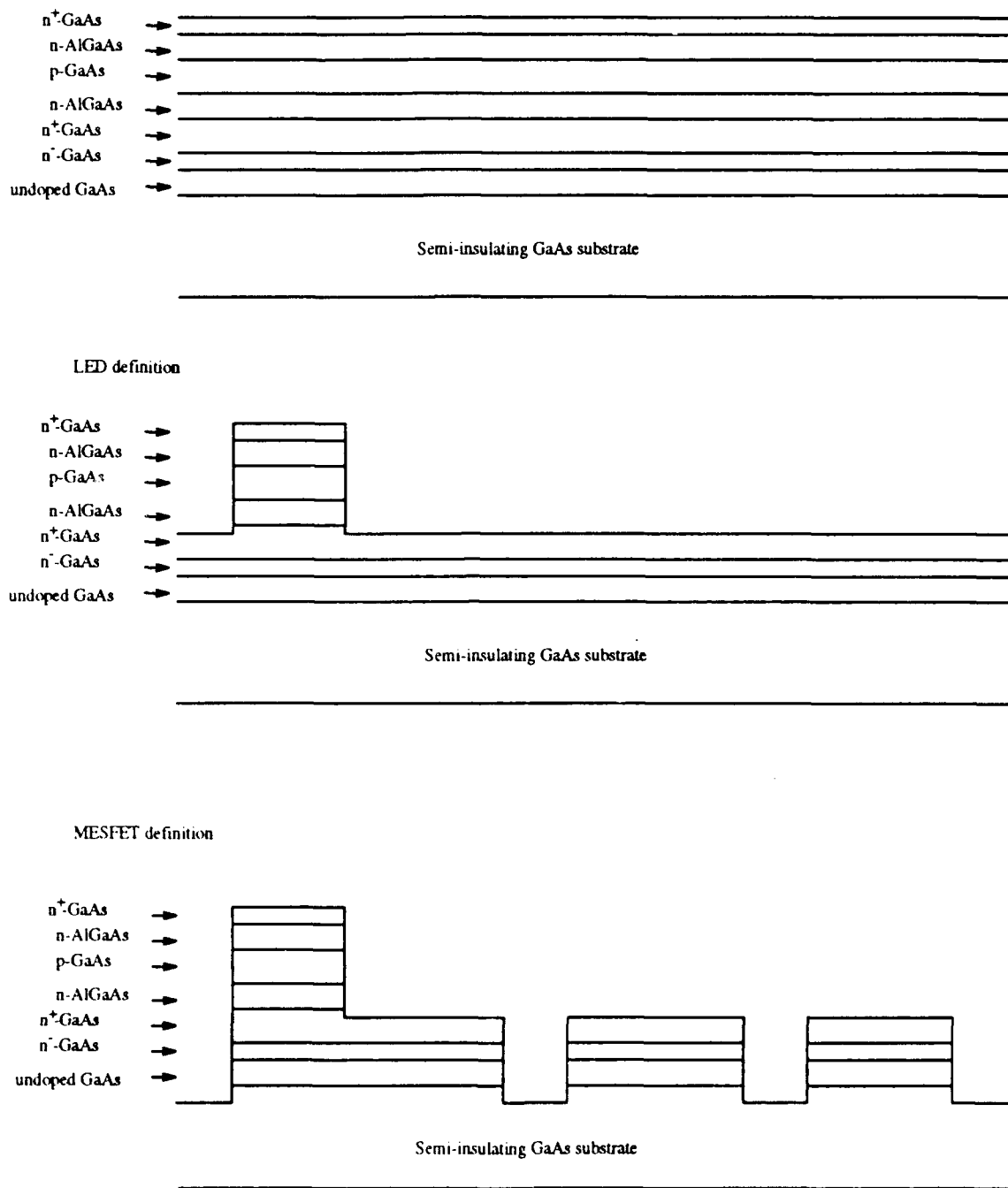

Fig. 3.25 Time response of the improved MESFET-based optoelectronic neuron. The rise time was measured to be 65 μsec.
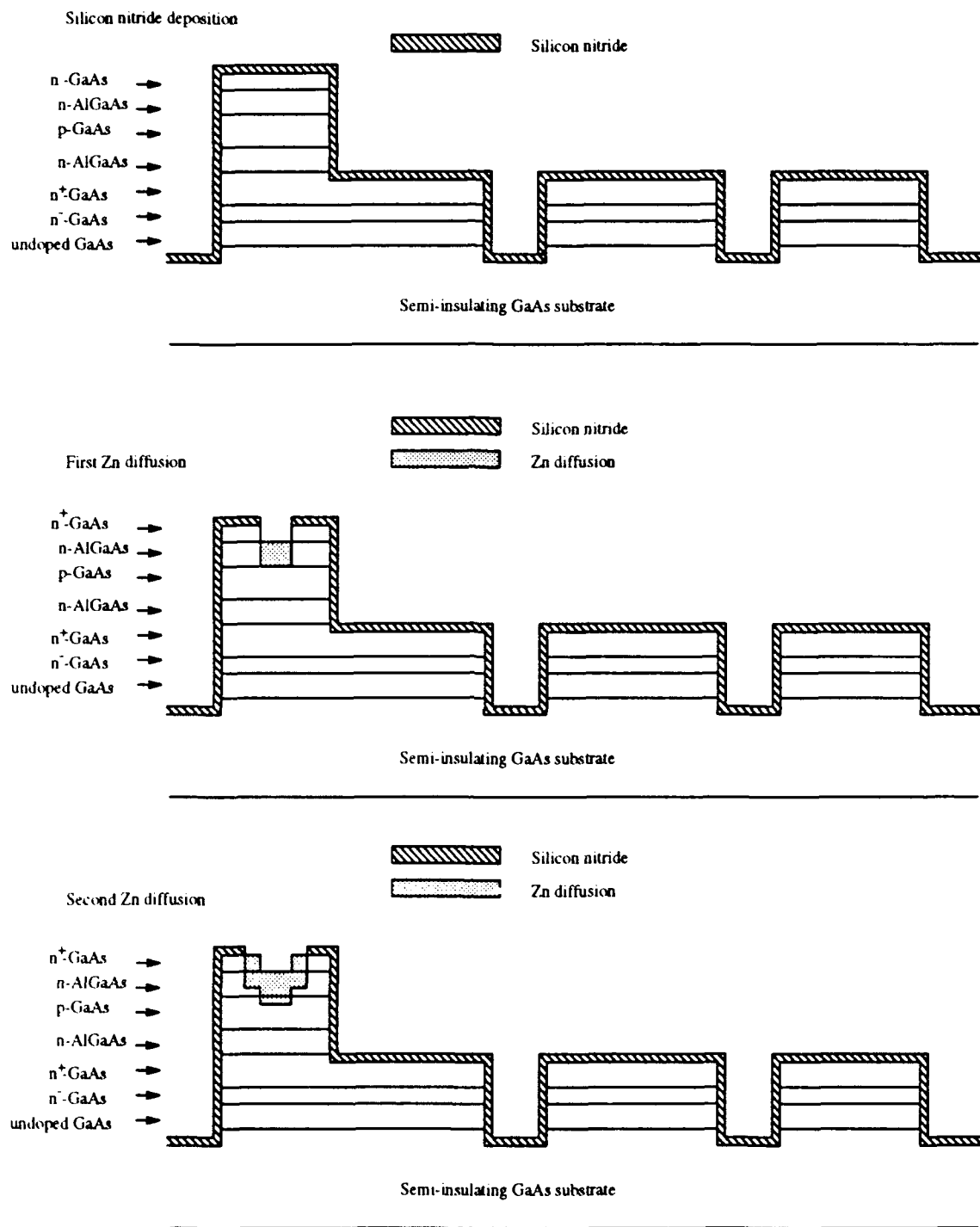
Because of the relatively thick base layer, the current gain, β, of the phototransistor was only about 3. This current gain dropped further as the input power level was reduced to the sub-μW regime. Since the overall goal of the MESFET-based optoelectronic neuron was to achieve a high-gain optical thresholding device at a low input power level, high-efficiency or even high-gain detectors at low input power level was vital to the success of the neuron. For this reason, optical FET's were developed. The operational principle of the optical FET was described in Sec. 3.3.3. In addition to the inherent high optical gain achievable in the optical FET, the structure of the optical FET was identical to that of a conventional MESFET. This meant that, for our MESFET-based neurons, optical FET's could be easily implemented into the existing material and process. This was a very important advantage of having the optical FET.

The fabricational steps of the MESFET-based neuron incorporating the optical FET as the detector were very similar to those of the conventional MESFET-based neuron. The difference was the definition of the optical FET rather than the phototransistor. Figure 3.26(a)-(d) show the sequential fabrication steps of the new neuron with Fig. 3.26(d) illustrating the entire device cross section of the neuron.

This neuron was tested at the same conditions as the previous one. Again, the gate of the loading MESFET in the input switching stage was left floating to minimize the gate leakage current. The optical input-output characteristics are shown in Fig. 3.27. Because of the insufficient recess in the gate of the output driving MESFET, this MESFET was not pinched off at zero gate bias. As a result, a current flowed between the source and the drain with zero input power onto the optical FET. This caused a non-zero LED output power at zero input power level. The remedy to this problem was to recess the gate of the LED-driving MESFET further until the current was close to zero at zero gate bias. This would shift the entire input-output curve shown in Fig. 3.27 down to the origin so that a normal

n+-GaAs →
n-AlGaAs →
p-GaAs →
n-AlGaAs →
n+-GaAs →
n--GaAs →
undoped GaAs →

Semi-insulating GaAs substrate

LED definition

n+-GaAs →
n-AlGaAs →
p-GaAs →
n-AlGaAs →
n+-GaAs →
n--GaAs →
undoped GaAs →

Semi-insulating GaAs substrate

MESFET definition

n+-GaAs →
n-AlGaAs →
p-GaAs →
n-AlGaAs →
n+-GaAs →
n--GaAs →
undoped GaAs →

Semi-insulating GaAs substrate

Fig. 3.26(a) Fabricational steps of the new MESFET-based optoelectronic neuron incorporating the optical FET as the detector.

Silicon nitride deposition

Silicon nitride

n -GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n⁺-GaAs
n⁻-GaAs
undoped GaAs

Semi-insulating GaAs substrate

First Zn diffusion

Silicon nitride

Zn diffusion

n⁺-GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
n⁺-GaAs
n⁻-GaAs
undoped GaAs

Semi-insulating GaAs substrate

Second Zn diffusion

Silicon nitride

Zn diffusion

n⁺-GaAs
n-AlGaAs
p-GaAs
n-AlGaAs
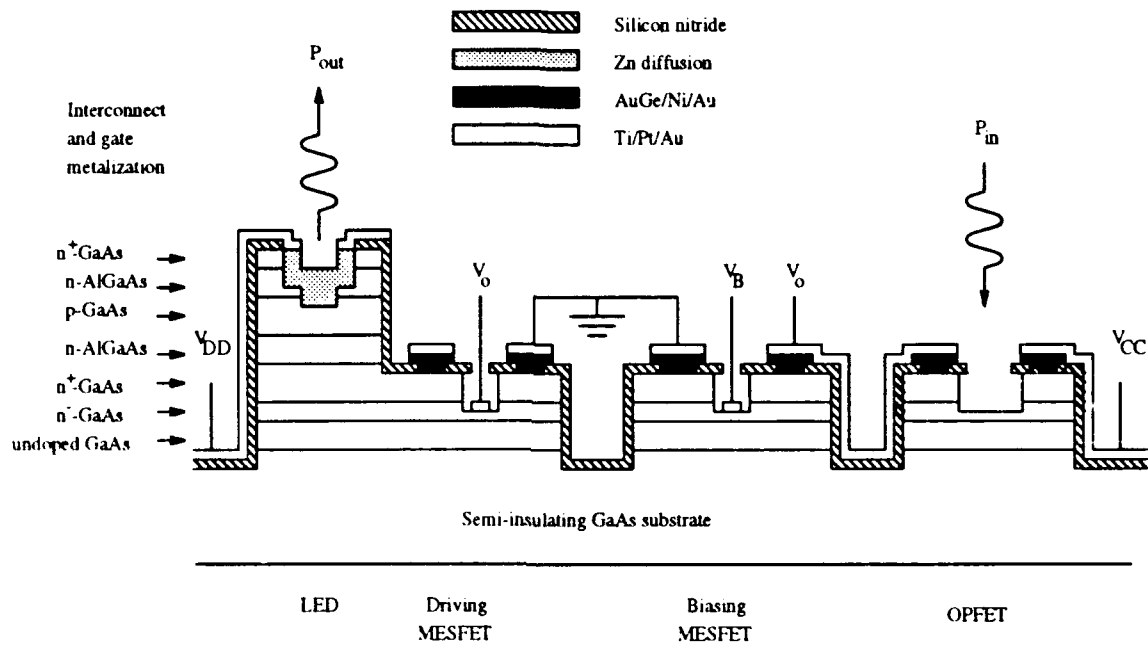n⁺-GaAs
n⁻-GaAs
undoped GaAs

Semi-insulating GaAs substrate
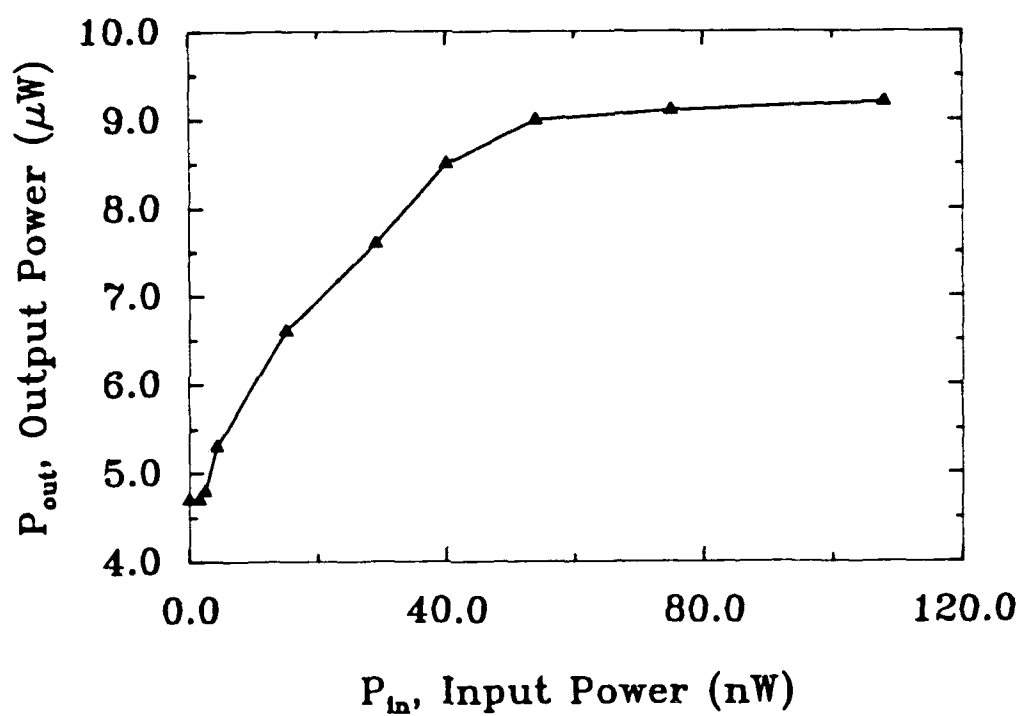
Fig. 3.26(b) Fabricational steps of the new MESFET-based optoelectronic neuron incorporating the optical FET as the detector.
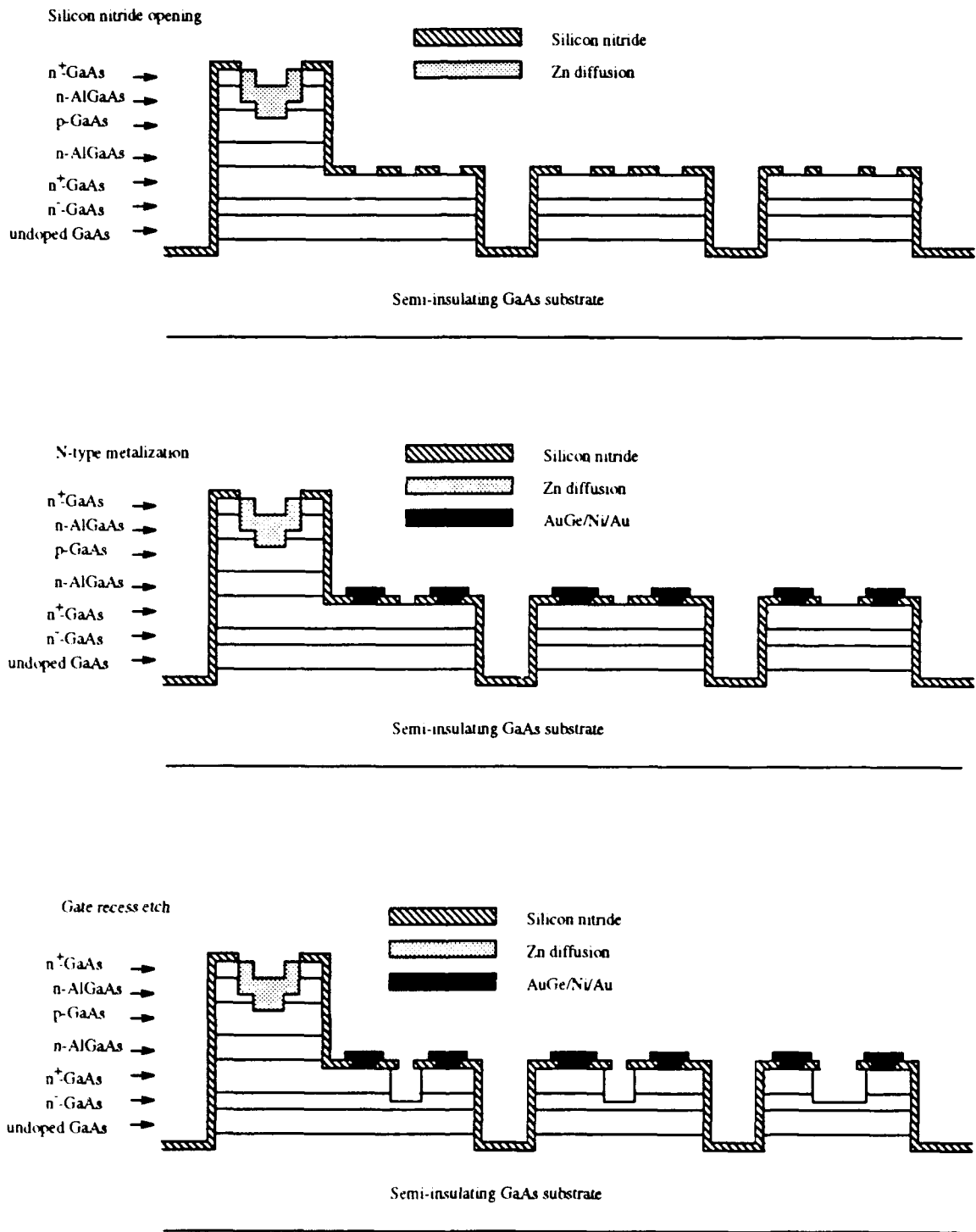
Fig. 3.26(c) Fabricational steps of the new MESFET-based optoelectronic neuron incorporating the optical FET as the detector.

Fig. 3.26(d) Complete device cross sectional view of the new MESFET-based optoelectronic neuron incorporating the optical FET as the detector.

Fig. 3.27 Input-output characteristics of the MESFET-based optoelectronic neuron that incorporated the optical FET in replacing the phototransistor. A differential optical gain of 80 was measured in this neuron.
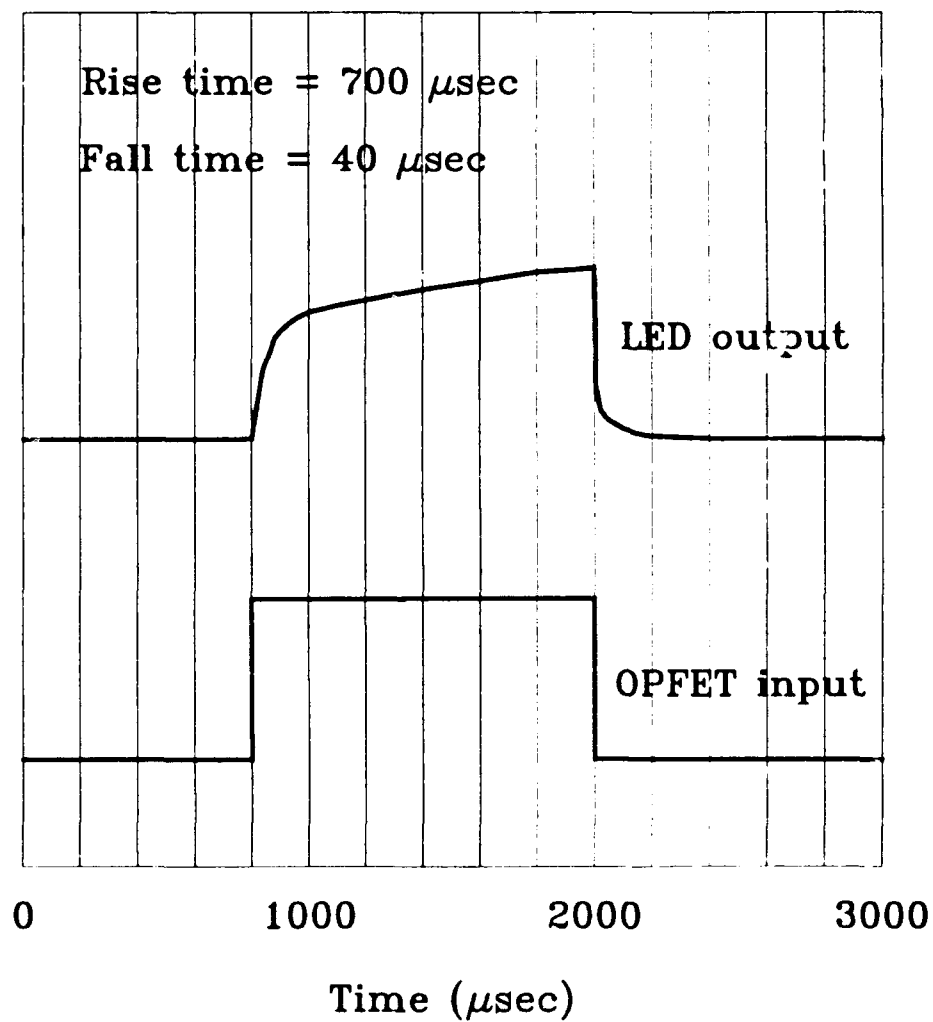
Fig. 3.28 Time response of the MESFET-based optoelectronic neuron incorporating the optical FET as the detector. The rise time was measured to 1 700 μsec.

neuron input-output characteristics could be obtained. Despite the gate recess problem, the differential optical gain measured was quite impressive. The output rose by 4.3 μW over an input swing of 54 nW. This corresponded to a differential optical gain of 80 [33]. It is also worth noting that the minimum input power needed to turn on the neuron had dropped significantly from the previous 1 μW down to about 5 nW. This could be attributed to the higher efficiency of the detector as well the overall reduction in the gate leakage current. Since this initial thresholding power was very small, the absolute optical gain was approximately the same as the differential optical gain, assuming the gate of the output MESFET was properly recessed. During the on-state of the neuron, the total current drawn by the LED was 0.9 mA, which implied an electrical power dissipation of 1.8 mW/neuron. Again, if the gate were properly recessed, this dissipation power would be reduced by 50%. The time response of the neuron was measured and is shown in Fig. 3.28. A rise time of 700 μsec was measured. When this was multiplied by the optical switching power of 54 nW, an optical switching energy of 38 pJ was obtained. Again, this was on the same order of magnitude as the previous optical switching energies. This indicated that the speed of the MESFET-based optoelectronic neurons was limited by the charging process of the gate capacitance and varied inversely proportional with the input power level. Table 3.1 summarizes the results of the MESFET-based optoelectronic presented in this report.

	Phototransistor-based neuron	Phototransistor-based neuron with gate leakage reduction	Optical FET-based neuron
OUTPUT POWER	12 μW	8 μW	9 μW
DIFFERENTIAL OPTICAL GAIN	6	40	80
ABSOLUTE OPTICAL GAIN	2.5	8	150
RISE TIME	5 μsec	65 μsec	700 μsec
SWITCHING POWER	2 μW	0.2 μW	54 nW
MINIMUM THRESHOLD	3 μW	1 μW	5 nW
OPTICAL SWITCHING ENERGY	10 pJ	13 pJ	38 pJ
ELECTRICAL POWER DISSIPATION	2.4 mW	1.6 mW	1.8 mW

Table 3.1 Summary of the neuron characteristics for three versions of MESFET-based optoelectronic neurons discussed in this report.

Reference

[1] F. H. Mok, M. C. Tackitt, and H. M. Stoll, "Storage of 500 High-Resolution Holograms in a LiNBO$_3$ Crystal," Opt. Lett., Vol. 16, pp. 605, 1991.

[2] D. Psaltis, D. Brady, and K Wagner, "Adaptive Optical Networks Using Photorefractive Crystals," Appl. Opt., Vol. 27, pp. 1752, 1988.

[3] E. G. Paek, and D. Psaltis, "Optical Associative Memory Using Fourier Transform Holograms," Opt. Eng., Vol 26, pp. 4?8, 1987.

[4] W. P. Bleha, L. T. Lipton, E. Wiener-Avnear, J. Grinberg, P. G. Reif, D. Cassasent, H. B. Brown, and B. V. Markevitch, "Application of the Liquid Crystal Light Valve to Real-Time Optical Data Processing," Opt. Eng., Vol. 17, pp. 371, 1978.

[5] L. K. Cotter, T. J. Drabik, R. J. Dillon, and M. A. Handschy, "Ferroelectric Liquid Crystal Silicon Integrated Circuit Spatial Light Modulator," Opt. Lett., Vol. 15, pp. 291, 1990.

[6] G. H. Haertling, "PLZT Electrooptic Materials and Applications - a Review," Ferroelectrics, Vol 75, pp. 25, 1987.

[7] H. K. Choi, G. W. Turner, J. C. C. Fan, J. M. Phillips, and B. Y. Tsaur, "Prospects for Monolithic GaAs/Si Integration," in Heteroepitaxy on Silicon II (J. C. C. Fan, J. M. Phillips, and B. Y. Tsaur eds.). Pittsburgh, PA : Material Research Society, pp. 213, 1987.

[8] D. R. Mayers, J. F. Klem, and J. A. Lott, "(AlGa)As/(InGa)As Strained-Quantum-Well FET's on Silicon Dioxide by Selective Device Lift-off as an Alternative to Heteroepitaxy," Tech. Dig. of International Electron Device Meeting, New York, pp. 704, 1988.

[9] N. Bar-Chaim, S. Margalit, A. Yariv, and I. Ury, "GaAs Integrated Optoelec-

tronics," IEEE Trans. Electron Dev., Vol. ED-29, pp. 1372, 1982.

[10] J. Katz, N. Bar-Chaim, P. C. Chen, S. Margalit, I. Ury, D. Wilt, M. Yust, and A. Yariv, "Monolithic Integration of a GaAlAs Buried-Heterostructure Laser and a Bipolar Phototransistor," Appl. Phys. Lett., vol. 37, pp. 211, 1980.

[11] H. Lin, and S. Lee, "Super-Gain AlGaAs/GaAs Heterojunction Bipolar Transistors Using an Emitter Edge-Thinning Design," Appl. Phys. Lett., Vol. 47, pp. 839, 1985.

[12] Y. Hasumi, A. Kozen, J. Temmyo, and H. Asahi, "A GaAs/GaAlAs Double-Heterojunction Device Function as a Bipolar Transistor and Injection Laser for Optoelectronic Integrated Circuits," IEEE Electron. Dev. Lett., vol. EDL-8, pp. 10, 1987.

[13] A. Cazarre, J. Tasselli, A. Marty, J. P. Baibe, and G. Rey, "GaAlAs/GaAs Heterojunction Bipolar Phototransistors Grown by LPE with a Current Gain of 50000," Electron. Lett., vol. 21, pp. 1124, 1985.

[14] A. Scavennec, D. Ankri, C. Besombes, C. Courbet, J. Riou, and F. Heliot, "High-Gain Low-Noise GaAlAs/GaAs Phototransistors," Electron. Lett., Vol. 19, pp. 394, 1983.

[15] H. Ito, T. Ishibashi, and T. Sugeta, "Fabrication and Characterization of Al-GaAs/GaAs Heterojunction Bipolar Transistors," IEEE Trans. Electron. Dev., Vol. 34, pp. 224, 1987.

[16] K Taira, C. Takano, H. Kawai, and M. Arai, "Emitter Grading in AlGaAs/ GaAs Heterojunction Bipolar Transistor Grown by Metalorganic Chemical Vapor Deposition," Appl. Phys. Lett., vol. 49, pp. 1278, 1986.

[17] J. H. Kim, S. H. Lin, J. Katz, and D. Psaltis, "Monolithically Integrated Two-Dimensional Arrays of Optoelectronic Threshold Devices for Neural Network Applications," SPIE Proc., vol. 1043-7, 1989.

[18] S. L. Su, O. Tejayadi, T. J. Drummond, R. Fisher, and H. Morkoc, "Double

Heterojunction Al_xGa_{1-x} As/GaAs Bipolar Transistors (DHBJT's) by MBE with a Current Gain of 1650," IEEE Electron. Dev. Lett., Vol. 4, pp. 130, 1983.

[19] D. Ankri, R. Azoulay, E. Caquot, J. Dangla, C. Dubon, and J. F. Palmier, "Analysis of d.c. Characterization of GaAlAs/GaAs Double Heterojunction Bipolar Transistors," Solid State Electron., Vol. 29, pp. 141, 1986.

[20] L. F. Eastman, and M. S. Shur, "Substrate Current in GaAs MESFET's," IEEE Trans. Electron. Dev., Vol. ED-26, pp. 1359, 1979.

[21] S. Sze, "Physics of Semiconductor Devices," pp. 312, John Wiley & Sons, Inc., 1981.

[22] R. A. Sadler, and L. F. Eastman, "High-Speed Logic at 300K with Self-Aligned Submicrometer-Gate GaAs MESFET's," IEEE Electron. Dev. Lett., Vol. EDL-4, pp. 215, 1983.

[23] H. M. Levy, and R. E. Lee, "Self-Aligned Submicron Gate Digital GaAs Integrated Circuits," IEEE Electron. Dev. Lett., Vol. EDL-4, pp. 102, 1983.

[24] Private Communication with Mr. David Forgerson of Vitesse Semiconductor Corp., Camarillo, CA.

[25] R. E. Williams, "Gallium Arsenide Processing Techniques," pp. 101, Artech House, Inc., 1984.

[26] H. M. Macksey, F. H. Doerbeck, and R. C. Vail, "Optimization of GaAs Power MESFET Device and Material Parameters for 15-GHz Operation," IEEE Trans. Electron. Devices, ED-27, pp. 467, 1980.

[27] T. Sugeta, and Y. Mizushima, "High Speed Photoresponse Mechanism of a GaAs-MESFET," Jpn. J. Appl. Phys., Vol. 19, pp. L27, 1980.

[28] J. C. Gummel, and J. M. Ballantyne, "The OPFET : A New High Speed Optical Detector," Technical Digest of the International Electron Device Meeting, pp. 120, 1978.

[29] J. C. Gummel, and J. M. Ballantyne, Comments on "High Speed Photoresponse Mechanism of a GaAs-MESFET," Jpn. J. Appl. Phys., Vol. 19, pp. L273, 1980.

[30] A. L. Lentine, H. S. Hinton, D. A. B. Miller, J. E. Henry, J. E. Cunningham, and L. M. F. Chirovsky, "Symmetric Self-Electrooptic Effect Device : Optical Set-Reset Latch, Differential Logic Gate, and Differential Modulator/Detector," IEEE J. Quantum Electron., Vol. QE-25, pp. 1928, 1989.

[31] S. H. Lin, F. Ho, J. H. Kim, and D. Psaltis, "GaAs-Based Optoelectronic Neurons," Technical Digest of Optical Computing Conference, Salt Lake City, UT, March 4-6, pp. 295, 1991.

[32] S. H. Lin, F. Ho, J. H. Kim, and D. Psaltis, "Monolithic Integrated Opto-electronic Thresholding Devices for Neural Network Applications," Technical Digest of Conference on Lasers and Electro-Optics, Baltimore, MD, May 12-17, pp. 82, 1991.

[33] S. H. Lin, D. Psaltis, and J. H. Kim, "High-Gain GaAs Optoelectronic Thresholding Devices for Optical Neural Network Applications," Technical Digest of Integrated Photonic Research Conference, Monterey, CA, April 9-11, pp.8, 1991.